TABLE 1.—REGULATIONS IMPLEMENTING THE HAZARDOUS AND SOLID WASTE AMENDMENTS OF 1984

Promulgation date	Title of regulation	Federal Register reference	Effective date
• •	• •	• •	•
(Publication date of final rule).....	Air standards for Tanks, Surface Impoundments, and Containers.	(Insert Federal Register reference of final rule).	(Publication date of final rule plus 6 months).

§ 271.1 [Amended]

44. Section 271.1(j) is amended by adding the following entry to Table 2 in chronological order by date of publication:

TABLE 2.—SELF-IMPLEMENTING PROVISIONS OF THE HAZARDOUS AND SOLID WASTE AMENDMENTS OF 1984

Effective date	Self-implementing provision	RCRA citation	Federal Register reference
• •	• •	• •	•
(Publication date of final rule plus 6 months).	Air standards for Tanks, Surface Impoundments, and Containers.	3004(n).....	(Insert Federal Register reference of final rule).

[FR Doc. 91–16416 Filed 7–19–91; 8:45 am]
BILLING CODE 6560–50–M

federal register

Monday
July 22, 1991

Part III

Department of Commerce

Office of the Secretary

Adjustment of the 1990 Census for Overcounts and Undercounts of Population and Housing; Notice of Final Decision

DEPARTMENT OF COMMERCE

Office of the Secretary

[Docket No. 91282-1181]

Decision of the Secretary of Commerce on Whether a Statistical Adjustment of the 1990 Census of Population and Housing Should Be Made for Coverage Deficiencies Resulting in an Overcount or Undercount of the Population

AGENCY: U.S. Department of Commerce.

ACTION: Notice of final decision.

SUMMARY: This is a notice of the final decision of the Secretary of Commerce on the issue of adjusting the 1990 census to correct for overcounts or undercounts of the population in the 1990 Decennial Census. The purpose of this notice is to inform the public of the decision and to explain the basis for the decision.

DATES: The decision is effective on July 15, 1991.

FOR FURTHER INFORMATION CONTACT: Michael R. Darby, Under Secretary for Economic Affairs and Administrator, Economics and Statistics Administration, Room 4848 Herbert C. Hoover Building, United States Department of Commerce, Washington, DC 20230, Telephone (202) 377-3727.

SUPPLEMENTARY INFORMATION: The Secretary of Commerce is required, pursuant to 13 U.S.C. 141, to conduct a decennial census of the population of the United States. The population totals derived from the census provide the basis for the apportionment of seats in the United States House of Representatives, for state legislative redistricting, for determining district boundaries for county and city elections, and for the allocation of federal funds to state and local governments.

In 1987, the Secretary of Commerce decided not to plan for a statistical adjustment of the 1990 census. As a result, a lawsuit was filed by the city of New York and other parties seeking to compel the Department to plan for such an adjustment. Pursuant to an agreement between the parties in *City of New York. et al.* v. *Department of Commerce. et al.*, 88-Civ.-3474 (E.D.N.Y.), the Department undertook a *de novo* review of the adjustment issue in order to make a decision no later than July 15, 1991, on whether to adjust the 1990 census. The purpose of this notice is to inform the public about the Secretary's decision and the basis for the decision.

Final guidelines which aided the Secretary in his decision were published in the **Federal Register** on March 15, 1990 (FR vol. 55, no. 51, part III pp. 9838-9861).[1] They were intended to provide the framework for a balanced consideration by the Secretary of factors relevant to the decision.

The census adjustment decision process was divided into several distinct phases. The first phase was the actual enumeration of the population. The second phase was the conduct of a post-enumeration survey, based on a probability sample of housing units. This sample provided data for two purposes: estimation of the net overcount or undercount of basic enumeration subgroups using capture-recapture methodology, and application of factors for the adjustment of the enumerated counts. The third phase of the process was a determination of the adequacy of the post-enumeration survey as an evaluation and adjustment tool. The fourth and final phase of the process was a decision on the adjustment question by the Secretary based on the published guidelines.

In making his decision, the Secretary relied on the advice of senior officials in the Economics and Statistics Administration, which includes the Census Bureau, as well as other senior advisors. The Secretary also relied on the individual recommendations of the eight members of the Special Advisory Panel appointed to provide independent advice to the Secretary on the adjustment question. In addition, the Secretary considered the public comments submitted to the Department pursuant to a **Federal Register** notice dated May 24, 1991, seeking comments on the question of whether the 1990 Census should be adjusted. The Department received approximately 650 public comments. These comments, as well as the appendices referred to in the following explanation of the decision, are available for public inspection in the U.S. Department of Commerce Central Reference and Records Inspection Facility, room 6020 Herbert C. Hoover Building, 14th Street and Constitution Avenue, NW., Washington, DC 20230.

Following is a detailed discussion of the adjustment decision and the basis for the decision. The discussion is in four sections: a summary statement, an analysis of the guidelines, an evaluation of the recommendations of the Special Advisory Panel and a statement of the decennial census procedures.

[1] Proposed guidelines were published in the **Federal Register** on December 11, 1989. The Court has previously considered and rejected a challenge to the guidelines. See *City of New York* v. *United States Department of Commerce*, 739 F.Supp. 767 (E.D.N.Y. 1990).

Dated: July 15, 1991.

Robert A. Mosbacher,

Secretary of Commerce.

SECTION 1—SUMMARY STATEMENT

Statement of Secretary Robert A. Mosbacher on Adjustment of the 1990 Census

Reaching a decision on the adjustment of the 1990 census has been among the most difficult decisions I have ever made. There are strong equity arguments both for and against adjustment. But most importantly, the census counts are the basis for the political representation of every American, in every state, county, city, and block across the country.

If we change the counts by a computerized, statistical process, we abandon a two hundred year tradition of how we actually count people. Before we take a step of that magnitude, we must be certain that it would make the census better and the distribution of the population more accurate. After a thorough review, I find the evidence in support of an adjustment to be inconclusive and unconvincing. Therefore, I have decided that the 1990 census counts should not be changed by a statistical adjustment.

The 1990 census is one of the two best censuses ever taken in this country. We located about 98 percent of all the people living in the United States as well as U.S. military personnel living overseas, which is an extraordinary feat given the size, diversity and mobility of our population. But I am sad to report that despite the most aggressive outreach program in our nation's history, census participation and coverage was lower than average among certain segments of our population. Based on our estimates, Blacks appear to have been undercounted in the 1990 census by 4.8%, Hispanics by 5.2%, Asian-Pacific Islanders by 3.1%, and American Indians by 5.0%, while non-Blacks appear to have been undercounted by 1.7%.

I am deeply troubled by this problem of differential participation and undercount of minorities, and I regret that an adjustment does not address this phenomenon without adversely affecting the integrity of the census. Ultimately, I had to make the decision which was fairest for all Americans.

The 1990 census is not the vehicle to address the equity concerns raised by the undercount. Nonetheless, I am today requesting that the Census Bureau incorporate, as appropriate, information gleaned from the Post-Enumeration Survey into its intercensal estimates of

the population. We should also seek other avenues for the Bush Administration and Congress to work together and address the impact of the differential undercount of minorities on federal programs.

In reaching the decision not to adjust the census, I have benefitted from frank and open discussions of the full range of issues with my staff, with senior professionals from the Economics and Statistics Administration and the Census Bureau, with my Inspector General, and with statisticians and other experts. Throughout these discussions, there was a wide range of professional opinion and honest disagreement. The Department has tried to make the process leading to this decision as open as possible. In that spirit, we will provide the full record of the basis for our decision as soon as it is available.

In reaching the decision, I looked to statistical science for the evidence on whether the adjusted estimates were more accurate than the census count. As I am not a statistician, I relied on the advice of the Director of the Census Bureau, the Associate Director for the Decennial Census and other career Bureau officials, and the Under Secretary for Economic Affairs and Administrator of the Economics and Statistics Administration. I was also fortunate to have the independent counsel of the eight members of my Special Advisory Panel. These eight experts and their dedicated staffs gave generously of their time and expertise, and I am grateful to them.

There was a diversity of opinion among my advisors. The Special Advisory Panel split evenly as to whether there was convincing evidence that the adjusted counts were more accurate. There was also disagreement among the professionals in the Commerce Department, which includes the Economics and Statistics Administration and the Census Bureau. This compounded the difficulty of the decision for me. Ultimately, I was compelled to conclude that we cannot proceed on unstable ground in such an important matter of public policy.

The experts have raised some fundamental questions about an adjustment. The Post-Enumeration Survey, which was designed to allow us to find people we had missed, also missed important segments of the population. The models used to infer populations across the nation depended heavily on assumptions, and the results changed in important ways when the assumptions changed. These problems don't disqualify the adjustment automatically—they mean we won't get a perfect count from an adjustment. The question is whether we will get better estimates of the population. But what does better mean?

First, we have to look at various levels of geography—whether the counts are better at national, state, local, and block levels. Secondly, we have to determine both whether the actual count is better and whether the share of states and cities within the total population is better. The paradox is that in attempting to make the actual count more accurate by an adjustment, we might be making the shares less accurate. The shares are very important because they determine how many congressional seats each state gets, how political representation is allocated within states, and how large a "slice of the pie" of federal funds goes to each city and state. Any upward adjustment of one share necessarily means a downward adjustment of another. Because there is a loser for every winner, we need solid ground to stand on in making any changes. I do not find solid enough ground to proceed with an adjustment.

To make comparisons between the accuracy of the census and the adjusted numbers, various types of statistical tests are used. There is general agreement that at the national level, the adjusted counts are better, though independent analysis shows that adjusted counts, too, suffer from serious flaws. Below the national level, however, the experts disagree with respect to the accuracy of the shares measured from an adjustment, The classical statistical tests of whether accuracy is improved by an adjustment at state and local levels show mixed results and depend critically on assessments of the amount of statistical variation in the survey. Some question the validity of these tests, and many believe more work is necessary before we are sure of the conclusions.

Based on the measurements so far completed, the Census Bureau estimated that the proportional share of about 29 states would be made more accurate and about 21 states would be made less accurate by adjustment. Looking at cities, the census appears more accurate in 11 of the 23 metropolitan areas with 500,000 or more persons: Phoenix, Washington, DC, Jacksonville, Chicago, Baltimore, New York City, Memphis, Dallas, El Paso, Houston and San Antonio. Many large cities would appear to be less accurately treated under an adjustment. While these analyses indicate that more people live in jurisdictions where the adjusted counts appear more accurate, one third of the population lives in areas where the census appears more accurate. As the population units get smaller, including small and medium sized cities, the adjusted figures become increasingly unreliable. When the Census Bureau made allowances for plausible estimates of factors not yet measured, these comparisons shifted toward favoring the accuracy of the census enumeration. Using this test, 28 or 29 states were estimated to be made less accurate if the adjustment were to be used. What all these tests show, and no one disputes, is that the adjusted figures for some localities will be an improvement and for others the census counts will be better. While we know that some will fare better and some will fare worse under an adjustment, we don't really know how much better or how much worse. If the scientists cannot agree on these issues, how can we expect the losing cities and states as well as the American public to accept this change?

The evidence also raises questions about the stability of adjustment procedures. To calculate a nationwide adjustment from the survey, a series of statistical models are used which depend on simplifying assumptions. Changes in these assumptions result in different population estimates. Consider the results of two possible adjustment methods that were released by the Census Bureau on June 13, 1991. The technical differences are small, but the differences in results are significant. The apportionment of the House of Representatives under the selected scheme moved two seats relative to the apportionment implied by the census, whereas the modified method moved only one seat. One expert found that among five reasonable alternative methods of calculating adjustments, none of the resulting apportionments of the House were the same, and eleven different states either lost or gained a seat in at least one of the five methods. I recognize that the formulas for apportioning the House are responsive to small changes and some sensitivity should be expected. What is unsettling, however, is that the choice of the adjustment method selected by Bureau officials can make a difference in apportionment, and the political outcome of that choice can be known in advance. I am confident that political considerations played no role in the Census Bureau's choice of an adjustment model for the 1990 census. I am deeply concerned, however, that adjustment would open the door to political tampering with the census in the future. The outcome of the enumeration process cannot be directly affected in such a way.

My concerns about adjustment are compounded by the problems an adjustment might cause in the redistricting process, which is contentious and litigious enough without an adjustment. An adjusted set of numbers will certainly disrupt the political process and may create paralysis in the states that are working on redistricting or have completed it. Some people claim that they will be denied their rightful political representation without an adjustment. Those claims assume that the distribution of the population is improved by an adjustment. This conclusion is not warranted based on the evidence available.

I also have serious concerns about the effect an adjustment might have on future censuses. I am worried that an adjustment would remove the incentive of states and localities to join in the effort to get a full and complete count. The Census Bureau relies heavily on the active support of state and local leaders to encourage census participation in their communities. Because census counts are the basis for political representation and federal funding allocations, communities have a vital interest in achieving the highest possible participation rates. If civic leaders and local officials believe that an adjustment will rectify the failures in the census, they will be hard pressed to justify putting census outreach programs above the many other needs clamoring for their limited resources. Without the partnership of states and cities in creating public awareness and a sense of involvement in the census, the result is likely to be a further decline in participation.

In looking at the record of public comment on this issue, I am struck by the fact that many civic leaders are under the mistaken impression that an adjustment will fix a particular problem they have identified—for example, specific housing units or group quarters that they believe we missed. It does not do so. It is not a recount. What an adjustment would do is add over 6 million unidentified people to the census by duplicating the records of people already counted in the census while subtracting over 900,000 people who were actually identified and counted. The decisions about which places gain people and which lose people are based on statistical conclusions drawn from the sample survey. The additions and deletions in any particular community are often based largely on data gathered from communities in other states.

The procedures that would be used to adjust the census are at the forefront of statistical methodology. Such research deserves and requires careful professional scrutiny before it is used to affect the allocation of political representation. Since the results of the evaluation studies of the survey were made available, several mistakes have been found which altered the certainty of some of the conclusions drawn by my advisors. The analysis continues, and new findings are likely. I am concerned that if an adjustment were made, it would be made on the basis of research conclusions that may well be reversed in the next several months.

It is important that research on this problem continue. We will also continue the open discussion of the quality of the census and the survey and will release additional data so that independent experts can analyze it. We must also look forward to the next census. Planning for the year 2000 has begun. A public advisory committee on the next census has been established and by early fall I will announce the membership of that committee. I have instructed the Census Bureau's Year 2000 task force to consider all options for the next census, including methods for achieving sound adjustment techniques.

I give my heartfelt thanks to the many people who have devoted so much time and energy to this enterprise. The staff at the Census Bureau have demonstrated their professionalism at every turn through the last two difficult years. They executed a fine census and an excellent survey and then condensed a challenging research program into a few short months. I am deeply grateful for their help. Let me reiterate my sincere thanks to the Special Advisory Panel for their substantial contribution. The staff at the Department, especially those in the Economics and Statistics Administration, also deserve praise.

With this difficult decision behind us, we will commit ourselves anew to finding sound, fair and acceptable ways to continue to improve the census process. We welcome the leadership of Congress and other public officials, community groups, and technical experts in maximizing the effectiveness and minimizing the difficulties of the year 2000 census.

July 15, 1991.

SECTION 2—ANALYSIS OF THE GUIDELINES

Analysis of the Guidelines

Introduction

The 1990 census counts should not be changed by a statistical adjustment. This section explains my evaluation of the evidence relevant to each of the eight guidelines that I considered in reaching my decision. Each section begins with a statement of the guideline and a reiteration of the explanation of the guideline contained in the March 15, 1990, **Federal Register** notice. A discussion of the guideline follows. The final section states my conclusions.

Summaries of my conclusions on each of the eight guidelines are set forth below.

Guideline One

Guideline One requires that convincing evidence be offered that the adjusted estimates of the population are more accurate than the census at the national, State, and local levels. In the absence of such evidence, the census counts are concluded to be the most accurate.

At the national level, it is likely that the PES-adjusted estimates reflect more accurately the total population and the racial and ethnic populations of the country. It appears equally clear, however, that the PES omitted large numbers of certain groups—notably black males. We have no information on the location of these persons. In addition, the PES and demographic analysis lead to sharply different conclusions about the accuracy of the census for several age/sex groups at the national level. Although these are not definitive disqualifiers at the national level, they do raise some question as to whether the adjusted figures are more accurate than the census count even at the national level.

The Constitution requires a census every 10 years not just to count the total number of people in the United States but to locate them so that political representation can be allocated to the states and the people in them in proportion to their numbers. I conclude that the primary criterion for accuracy should be distributive accuracy—that is, getting most nearly correct the proportions of people in different areas. Improved numeric accuracy, although in itself desirable, cannot compensate for treating states and individuals less fairly.

At the State and local level the correct statistical analysis for both distributive and numeric accuracy simply has not been completed. The total error model indicates that the adjusted figures tend to be too high but generally closer in numeric terms to the true population than the census counts which tend to be too low. However, there is sufficient uncertainty about the true variance of the adjusted figures that even numeric accuracy has not been definitively

demonstrated. The loss function analysis and hypothesis tests that have been prepared by the Census Bureau to date, although of uncertain reliability, do support the superior accuracy of the census counts versus the adjusted figures when we consider distributive accuracy—or fairness—and use reasonable estimates of the error variance of the alternative DSE. That is, for the Constitutional purposes of the census the available evidence is consistent with the census counts being more accurate than the adjusted counts. There is certainly not sufficient evidence to reject the distributive accuracy of the census counts in favor of the adjusted counts.

I conclude that, in accordance with Guideline One, the census counts are the most accurate count of the population of the United States at the State and local levels. While the preponderance of the evidence leads me to believe that the total population at the national level falls between the census counts and the adjusted figures, that conclusion is not relevant to the determination of distributive accuracy. Thus this guideline weighs in favor of a decision not to adjust.

Guideline Two

I conclude that the considerations pointed to by Guideline Two tend to reject use of the adjusted figures and support use of the census counts. The adjusted figures—like the census counts—are consistent across all jurisdictional levels and of sufficient detail for all purposes. However, the adjusted figures do not appear to be of sufficient quality to be usable for reapportionment and redistricting. First, the distributive accuracy of the census counts is superior as concluded above in my review of the evidence on Guideline One. Furthermore, substantial evidence casts doubt on the homogeneity assumption underlying the entire synthetic adjustment methodology. Even if the tests discussed under Guideline One and based on the homogeneity assumption had proven favorable to adjustment, this evidence would weigh against adjustment. Instead, both considerations imply that the adjusted figures are not of sufficient quality to be usable for reapportionment and legislative redistricting. Thus, this Guideline weighs in favor of a decision not to adjust the census.

Guideline Three

I have previously concluded that the adjusted figures have not been shown to be more accurate than the census enumeration. That is all that is required under Guideline Three to conclude that the census may not be adjusted. There are, however, additional considerations under Guideline Three under which I also conclude the 1990 census should not be adjusted.

It has proved virtually an impossible task to prespecify the adjustment procedure. It is equally impossible to prespecify the Census procedure. However, in the adjustment procedure an individual or responsible group must make choices which have politically significant effects on the counts that can be transparent to those making the choices. This puts the counts at greater risk of being manipulated than the census. There is no evidence of unprofessional or political manipulation in the 1990 PES program.

The results of the adjustment procedure are broadly robust at an aggregate, national level. However, although various alternatives seem to distribute counts in roughly similar ways, small changes in methodology can move seats in the House. It is also true that small changes in the census enumeration can move seats in the House as well, but no individual involved in the enumeration process can predict how. That is not true for the decisions for adjustment that cannot be or were not prespecified.

One of the most problematic parts of the adjusted process was the bundle of statistical techniques contained in the smoothing process. These techniques relied heavily on statistical assumptions, resulted in large changes in adjustment factors, and may very well have led to an overstatement of the undercount. Thus, this guideline weighs in favor of a decision not to adjust.

Guideline Four

Based on the information available, I conclude that an adjustment would adversely affect future census efforts to a greater extent than any adverse effects of a decision not to adjust. The evidence indicates that the controversy over adjustment is likely to have a negative effect on future censuses regardless of the outcome of the adjustment decision. I am concerned that an adjustment would reduce state and local support for future censuses, adversely affect the Department's ability to obtain appropriate funding for future censuses, adversely affect the quality of the work done in the future by temporary census enumerators who are essential in reaching the hard-to-count, subject the Census Bureau to partisan pressures, and create the possibility for political manipulation of future census counts. Thus, this guideline weighs in favor of a decision not to adjust.

Guideline Five

The question whether the chosen method of adjustment would violate the Constitution and federal statutes depends upon the substantive analysis of whether accuracy of the census is improved by an adjustment. Because there are other compelling substantive reasons not to adjust, legal considerations did not provide a basis for my decision.

Guideline Six

An adjustment to the census is a fundamental change in the way we count and locate the persons residing in the United States. I am deeply concerned that if an adjustment is made, it would be made on the basis of research conclusions that may very well be reversed in the next several months. That would be bad for the country and bad for the Census Bureau.

The results of the PES evaluation studies are not yet completely analyzed. Because of the compressed time schedule imposed by the July 15 deadline, the analysis has not been subject to the full professional scrutiny that such important research requires and deserves. To the Census Bureau's great credit, the statistical tools used to calculate and evaluate the adjusted counts are at the cutting edge of statistical research. But such cutting edge research is not tried and true—it requires more thorough scrutiny before it can be used to affect the allocation of political representation and Federal funding.

Nonetheless, the demands of good research must be weighed against the need for a timely decision. In time we may find a way of combining the PES and the census to create counts that better reflect the absolute levels and the distribution of the population. There are sufficient data and analysis to support a decision not to adjust.

Guideline Seven

Any decision will result in some level of disruption through legal challenges. On balance, the record indicates that a decision to adjust would likely be more disruptive than a decision not to adjust. A decision to adjust would clearly cause disruption in those States that have early redistricting deadlines. The assertion that persons are denied their rightful claims without an adjustment assumes that the distribution of the population is improved by an adjustment. Based on the evidence, this assumption is invalid. Thus, this guideline weighs in favor of a decision not to adjust.

Guideline Eight

The requirements for this Guideline have been met. This Guideline does not weigh in favor of a decision either way since the requirements of this Guideline could have been fully met if the decision had been to adjust.

Guideline One

The Census shall be considered the most accurate count of the population of the United States, at the national, State and local level, unless an adjusted count is shown to be more accurate. The criteria for accuracy shall follow accepted statistical practice and shall require the highest level of professional judgment from the Bureau of the Census. No statistical or inferential procedure may be used as a substitute for the Census. Such procedures may only be used as supplements to the Census.

Explanation

The mandate of the Census Bureau is to enumerate the population in a manner that assures that the count of the population is the best achievable given current methodology. As stated in the introduction, the assertion that a method involving statistical inference could lead to a more accurate enumeration warrants close scrutiny.

A set of adjusted counts would be based on a statistical inference that unaccounted for persons were present and that persons who were actually enumerated do not exist or were counted twice. Both determinations are based on a survey of a sample of similar blocks from locations across the country. Thus, the evidence, to be acceptable, must show convincingly that the count can be improved by statistical adjustment at national, state and local levels. In making this assessment, we will examine the effects of the proposed adjustment on the accuracy of counts at all geographic levels.

Comparison of estimates of population size. The estimates of the size of the population from the original enumeration, the demographic analysis, and the post-enumeration-survey estimates will be compared to assess their consistency. The comparison will take into consideration the uncertainty inherent in the demographic analysis and post-enumeration-survey estimates. For the reasons explained in the introduction, the original enumerations will be considered to be more accurate for all geographic areas unless the evidence from demographic analysis and the post-enumeration survey demonstrates convincingly that the dual-system estimate is more accurate.

Accordingly, the Bureau of the Census shall carefully scrutinize and fully describe the size of any net undercount or net overcount inferred from demographic analyses of population sub-groups and the sources of any net undercount or net overcount of population subgroups inferred from the analysis of the post-enumeration survey.

Technical Grounds

Demographic Analysis. Estimates of the size of certain cohorts of the population are based on assumptions about or studies of the behavior of those populations. For some cohorts these assumptions have led to conclusions of net undercounts or net overcounts in several different censuses. The extent to which such conclusions result from specific assumptions will be described. Moreover, the extent to which these assumptions are warranted, and the sensitivity of such conclusions to changes in these assumptions, will be assessed.

The potential sources of error in the demographic analyses the Bureau currently plans are:

Birth registration completeness.
Net immigration of undocumented aliens.
White births, 1915–1935.
Black births, 1915–1935.
Foreign-born emigrants.
Population over age 65.
Models to translate historical birth-record racial classifications into 1990 self-reported census concepts.

The Bureau will examine the effect of errors in each of these measurements on estimates of the net overcount or net undercount. These studies will yield ranges of uncertainty for the demographic estimates of the population which will in turn yield ranges of uncertainty for the net overcount or net undercount. The effect of uncertainty in each of these components will be cumulated into overall levels of potential error.

Post-Enumeration Survey. The capture-recapture method lies at the heart of the post-enumeration-survey models for estimating population coverage deficiencies. The use of this methodology to derive the net undercount or net overcount estimates will be clearly explained. The appropriateness of this methodology to the enumeration of the population will be assessed.

Like demographic analysis, the post-enumeration-survey adjustment mechanism relies on numerous assumptions. The extent to which these assumptions are warranted, and the sensitivity of the conclusions to changes in these assumptions, will be assessed.

Survey methods are based on randomly chosen samples that use statistical inference to estimate the population of the Nation and its components. Such estimates are subject to statistical variation within some range of values—that is, a replication of the process used to make the estimate (including taking the sample) may not lead to the same estimate as the original procedures. Thus, there is a likely range of estimates around the "true" count of the population that depends on the random sample chosen.

If the range of estimates likely to occur is small and near the "truth," then any particular estimate is close to the truth and, thus, acceptable as an approximation of the "truth." If the range is very large, then any particular estimate may not be close to the "truth," and the estimation process gives us little information about the "truth."

A relevant technical criterion related to uncertainty introduced by sampling is how small any possible range of dual-system estimates must be to conclude that any particular outcome of the dual-system estimation process is more accurate than the enumeration itself.

Because the post-enumeration survey itself is a sample, the quantified parameters of the deficiencies are themselves estimates and subject to statistical variability. This variability must be small enough to ensure that any modification of the enumeration is an improvement over the unadjusted counts.

The post-enumeration survey serves two functions. The first function is to detect any deficiencies in the enumeration. For the post-enumeration survey to show convincingly that the enumeration is deficient, it must be clear that the deficiencies are not a result of problems in taking the post-enumeration survey. It follows, then, that the quality of the post-enumeration survey is a central concern in the decision whether to adjust.

The second function is to quantify any deficiencies attributed to the enumeration precisely enough to allow the enumeration to be modified in such a way that we are reasonably certain that the modified enumeration is more accurate than the original enumeration. Thus the post-enumeration survey must quantify the deficiencies of the enumeration precisely and accurately.

How much uncertainty in the measures of deficiency of the enumeration is acceptable?

(1) If the likely range of measures of deficiency would include outcomes that would call for no modification in the

enumeration, then no modification would be done.

(2) The enumeration could be modified if the likely range of measures of deficiency would lead to potential modifications that would be substantially similar in terms of their impact on the counts of demographic groups, their impact on apportionment of Congress, and their impact on local population counts.

The quality of the net overcount or net undercount estimates that result from the post-enumeration survey depends on the quality of a series of operations used to gather and process the required data. The Bureau of the Census will undertake a series of studies to assess the statistical quality of the post-enumeration survey data. The results of these studies will yield measures of the precision and accuracy of the net overcount and net undercount estimates and a range of estimates for the net undercount and net overcount.

The current plans of the Bureau include investigation of the following sources of error for the dual system estimate of population size based on the post-enumeration survey and the census:

Missing data
Quality of the reported census day address
Fabrication in the P sample
Matching error
Measurement of erroneous enumerations
Balancing the estimates of gross overcount and gross undercount
Correlation bias
Random error

These and other component errors will be combined to produce an estimate of the overall level of error. In all evaluations, analyses will examine data for the population as a whole and for race, sex, Hispanic origin, and geographical detail.

Discussion

To certify a set of adjusted counts as the official counts of the population of the United States, one must accept the statistical inferences from a survey that there are persons who were unaccounted for by the census but who were actually present in a specific location on census day, that persons who were actually enumerated either did not exist or were counted twice, and that the same survey, when combined with census counts, can produce more accurate figures than the census enumeration alone. All these inferences are based on information from a sample of 377,381 persons in 171,390 housing units and group quarters in 5,290 block clusters. The people who are inferred to be missing from the census or erroneously enumerated in the census must then be correctly allocated to the specific blocks in which these mistakes were made. These blocks must be chosen out of the 4,830,514 inhabited blocks in the United States. Thus, acceptance of adjusted counts as more accurate requires not only that the counts themselves be shown to be more accurate, but that the distribution of those counts across the United States reflect more accurately the distribution of the population. This is the burden of proof imposed by Guideline One on any decision to adjust the census.

There are three population measurement techniques that play a role in making these statistical inferences. The first is the census enumeration. This was an effort to count each and every person residing in the United States on April 1, 1990. The second is the Post-Enumeration Survey (PES). This is the survey mentioned in the preceding paragraph that was taken several months after census day, independently of the census. An attempt is made to match the persons surveyed in the PES back to records in the census and to match persons in the census to the PES. From the results of this matching process, and a complex web of statistical models, inferences can be made about the number of persons missed by the census and their location. It is the quality of these inferences that is at issue. The third technique is called demographic analysis (DA). DA makes an independent estimate of the population at a national level from administrative records. It can be used to calibrate the results from the census or PES. DA calculates the population from the number of births, number of deaths, the number of immigrants, and the number of emigrants. It builds up a count of the population of the United States from birth and death certificates, immigration records and other sources. Like the census and the PES, DA is also an imperfect measure, so the quality of the inferences made from it are in question as well.

In the course of the discussion of this guideline, various aspects of these three complex processes will be explained and discussed. A detailed explanation can be found in Section Four of this report. We begin by comparing the national counts found in 1990 using these three methods.

A Comparison of the Counts at the National Level Using Three Methods

The national total count from the census enumeration is compared, in Table 1, Appendix 14, with the corresponding total in the proposed adjusted counts based on the PES and also with the corresponding estimates based on DA. The census count is 2.07% or 5,269,917 persons less than the PES estimate. There is evidence of racial, ethnic, and sex differential undercounts in the census when compared to the PES-based estimates. The count of black males in the census was 5.37% or 804,233 persons lower than the population inferred from the PES. The count of black females in the census was 4.33% or 715,543 persons lower than the PES estimated. For non-black males the census count was below the PES estimate by 2.02% or 2,205,443 persons and for non-black females the differential was 1.36% or 1,544,050 persons.

Estimates of national population totals are derived by DA based primarily on administrative records. Demographic analysis estimates provide national totals only and cannot be used to locate people as census counts are required to do. Many argue that the DA estimates broadly corroborate differential undercounts implied by PES-adjusted counts; [1] however, like the minority on the Undercount Steering Committee,[2] I find there are some important and puzzling differences. First, the overall undercount rate inferred from comparing the census to DA (1.85%) is smaller than that inferred from the PES (2.07%). At an aggregate level, the demographic analysis is thought to be more inclusive since the PES and census will miss people who are difficult to survey. Thus the estimate of the population from the PES was expected to be lower than the DA estimate. It is not. The PES estimated total population is 0.23% higher than the DA estimate. More detailed analysis shows that the PES and DA estimates are not far apart in a statistical sense.[3]

---

[1] See appendix 7: Bryant, Barbara E., Director of the U.S. Bureau of the Census, "Recommendation to Secretary of Commerce Robert A. Mosbacher on Whether or Not to Adjust the 1990 Census," June 28, 1991, [hereafter Bryant] page 16. See also Appendix 4, "Report of the Undercount Steering Committee," U.S. Bureau of the Census, June 21, 1991, [hereafter Undercount Steering Committee] page 4. See also Appendix 3: Ericksen, Eugene P., Estrada, Leobardo F., Tukey, John W., Wolter, Kirk M. "Report on the 1990 Decennial Census and the Post-Enumeration Survey," Members of the Special Advisory Panel, June 21, 1991, [hereafter Ericksen, *et al.*,] page 10.

[2] Undercount Steering Committee, page 4.

[3] The 95% confidence interval for the overall PES undercount rate is from 1.23% to 2.20% and the judgmental 95% confidence interval for the overall demographic undercount rate is from 1.6% to 3.4%. A confidence interval gives the range of statistically plausible values. The "95%" refers to the notion that one is 95% sure this interval has captured the true, but unknown, value. See table 2 in appendix 14.

Nevertheless, the fact that the direction of difference is the opposite of what statistical experience would have led us to expect raises a troubling question about the relationship between the two methods.[4]

Another example of a gross inconsistency between the PES and DA is that an adjustment would add 1,055,826 more females than DA indicates should be added. If DA were in fact correct, and the enumeration were adjusted, the official population counts would have a 0.82% overcount of females imbedded in it.

The third disturbing comparison between the PES and DA undercount rates is that all groups of black males (except those aged 10–19) are substantially undercovered by the PES relative to DA. This results in PES-based undercount rates that are substantially smaller than the DA rates. This is the type of result that is usually expected in comparing the PES and DA.[5] An adjustment based on the PES would add 804,233 black males to the population. According to demographic analysis, the number of black males that should be added to the population is 1,338,380. Thus the PES-based adjustment would be omitting 534,147 black males according to DA. For black females the PES adjustment would add 29,390 fewer persons than DA indicates should be added. If we accept the DA as being closer to the truth, we could not appropriately add the persons the PES missed to the count because we have no way of locating them.

Some will argue that "going part way" toward remedying the undercount of black males is better than doing nothing.[6] The trouble with this argument is that it ignores the fact that increased accuracy for census counts means not only increased accuracy in the level of the population, but also increased accuracy in the distribution of the population in states and localities. In particular, for the primary uses of the census—apportionment and redistricting—the share or fraction of the total population in a given state, city or precinct is critical. It is this fraction that determines political representation and the amount of Federal funds allocated across political jurisdictions. The paradox is that even if you improve the accuracy in the level of the population in any given city by adding at least some of the people missed in the census, you do not necessarily improve and can worsen accuracy in the share of the population in that city. This point is explored further in the section on how accuracy is measured.

Special Advisory Panel Member Wachter estimates that the number of people missed by both the census and the PES may be as high as half-a-million.[7] We do not know where these people are.[8] The implicit assumption that we would be making if we went ahead and adjusted the count is that they are spread over the country in the same way as the post-adjustment population. Such an assumption has no empirical foundation. There is no doubt that there is a fundamental deficiency in the count, but there is also a fundamental deficiency in the PES. It is not clear that the adjusted counts will accurately reflect relative populations in particular jurisdictions. As Wachter states:

> When we try to gauge the relative sizes of two states or cities or counties or districts [after an adjustment], we must always worry that there are enough more of the unreached in one than in the other to reverse the judgment about relative size that the adjusted counts would lead us to make.[9]

To further complicate matters, at the national level there are instances where a PES-based adjustment to the census would move subpopulation totals in the opposite direction from that indicated by DA:

- An adjustment based on the PES will add 180,318 non-black males aged 10–19, while the DA indicates 136,908 should be deleted, a difference in the wrong direction of 317,226.[10]
- An adjustment based on the PES will delete 91,631 males over the age of 65, while DA indicates that 192,950 should be added, a difference in the wrong direction of 284,541 persons.[11]
- An adjustment based on the PES will add 375,053 females aged 10–19 when DA indicates that 7,141 should be deleted, a difference of in the wrong direction of 382,191.[12]
- An adjustment based on the PES will delete 245,253 females over the age of 45 while DA indicates 146,255 should be added, a difference of 391,508 persons in the wrong direction.[13]

Another grouping of the population that plays a key role in the adjustment process is called a post-stratum. To calculate the adjusted population estimates, the population is broken down into 1392 groups called post-strata. Every individual in the United States fits into one, and only one, of these post-strata. These post-strata are based on census division, type of place of residence, tenure of residence, race, Hispanic ethnicity, sex, and age. These are the smallest groupings of people for which an undercount rate is estimated by the Census Bureau. When post-strata for similar types of persons are combined (for example, all post-strata with blacks, or all post-strata for people age 30–44) the results are largely consistent with expectations.[14] However, there is a lot of variation across the post-strata for similar types of people. Wachter offers intriguing evidence that "the story of census coverage, at a level of fine detail, is more complicated than one would hope." [15] For example, if one looks at all the post-strata for blacks, 25% of them show an overcount rather than an undercount.[16] Thus the broad, national-level aggregations of undercount by race, ethnicity, sex, and age mask a large amount of diversity within those groups. It is therefore overly simplistic to conclude that the census generally results in an undercount for all members of any particular group.

---

[4] As will be discussed later, there are measured biases in the production adjustment estimates. When corrections are made for these measured biases, the overall undercount rate measured by the PES falls below that of DA.

[5] The technical term for this is correlation bias.

[6] See Undercount Steering Committee, page 4; See also appendix 3: Ericksen, Eugene P. "Recommendation on 1990 Census Adjustment," Member, Special Advisory Panel, June 21, 1991, [hereafter Ericksen] page 2; See also Appendix 3: Estrada, Leobardo F. "Recommendation on 1990 Census Adjustment," Member, Special Advisory Panel, June 21, 1991, [hereafter Estrada] page 14; See also Appendix 3: Wolter, Kirk W. "Recommendation on 1990 Census Adjustment," Member, Special Advisory Panel, June 21, 1991, [hereafter Wolter] page 4.

[7] See appendix 3: Wachter, Kenneth W. "Recommendations on 1990 Census Adjustment," Member, Special Advisory Panel, June 17, 1991, [hereafter Wachter] page 8.

[8] The implications of this for accuracy are explained at length below.

[9] Wachter, page 8.

[10] The third table in appendix 14 shows that the 95% PES confidence interval for the undercount rate for this group is (0.53, 1.85) with a point estimate of 1.19. Demographic analysis shows a confidence range of [−1.21, 0.65) with a point estimate of −0.92. Thus neither estimate falls in the other's confidence range.

[11] The third and fourth table of appendix 14 show the confidence intervals for undercount rates for blacks and non-blacks separately. For non-blacks in this group, the confidence intervals for the two methods do not intersect, with the PES confidence interval completely less than zero and the DA confidence interval completely greater than zero. For blacks as well, the two intervals do not overlap. The PES spans zero, the DA is completely greater than zero.

[12] In appendix 14 the confidence intervals for this group are given for blacks and nonblacks separately. For non-blacks the intervals for the PES and DA do not overlap. For blacks they do.

[13] The confidence intervals for the four component groups are given in tables 1 and 2 of appendix 14. The intervals are wide enough that the differences may not be statistically significant.

[14] Wachter, pages 9–10.

[15] Wachter, page 10.

[16] Wachter, page 10.

This section has given an aggregate picture of the population using three different measurement instruments—the census, the PES, and DA. It is clear that the census suffers from an undercount, that the undercount is differential across race, ethnicity, and age, but that there is diversity within these groups. There are substantial and statistically different pictures of the population that are drawn by these three methods even at the national level. This is worrisome in and of itself. An adjustment based on the PES will be at face-value substantially different from our demographic estimates at the most aggregate levels. Whether it is an improvement depends not on its ability to add people and to subtract people from the census, but, rather, on its ability to add them and subtract them from the right places.

The Quality of the Census Enumeration

Special Advisory Panel Members Ericksen, Estrada, Tukey and Wolter all condemn the census as being fatally flawed.[17] I concede the census' imperfections, but the critical inquiry under this guideline is not how flawed the census is, but whether the PES can fix it.[18] Census taking is a complex task that must be completed within a short period of time. In an operation employing 350,000 temporary workers spread over more than 400 offices across the country, quality control is a real problem. The management information system the Census Bureau installed allowed the Census Bureau and the panelists to have access to the type of data panelists report. Thus, while identifying the flaws in the census is important for planning the next one, it simply begs the question that Guideline One poses: Is there convincing evidence showing that the adjustment is more accurate than the enumeration?

The Quality of the Alternative Measurement Tools

In considering whether to adjust the population for undercounts, the quality of the tools used to measure and then make an adjustment is important. The two methods that are alternative to the census are DA and the PES.

Demographic Analysis

Demographic analysis is a count of the aggregate population that is not based directly on any census. Instead it is built from administrative records including birth and death certificates, immigration records, and medicare records, among others. Limitations in record-keeping limit demographic breakdowns to those by age, sex, and black/non-black. There is no uniform reporting of ethnicity (*e.g.*, Hispanic origin) or the race of children of biracial couples. Even the same person might be reported as having different characteristics on birth and death records. Because we do not keep records of movements of individuals within the United States this analysis can only be done at the national level.

Furthermore, demographic estimates of the population are continually being changed. No demographic estimates are ever final, as new sources of data and statistical models are used presumably to improve the inferences made about the population. (For example, as a result of the demographic analysis for this census, the estimates of the 1980 population were still being changed as late as last month.) This year the Census Bureau undertook a series of investigations into the quality of the demographic estimates. An important improvement in the estimates was the first attempt to characterize the uncertainty inherent in them with uncertainty intervals about the point estimates. These improvements are reported in the demographic reports D1–D11.[19] Because demographic analysis will not be used in any adjustment, any detailed discussion of its results is foregone. Nevertheless, it is worth noting that the uncertainty intervals have been used in the previous descriptions of consistency of the various estimates of the population.

In an article in *Science*, David Freedman, Professor of Statistics at the University of California at Berkeley, discusses the limitations of DA in some detail.[20] Racial classification procedures vary widely. Incomplete coverage of vital statistics is a problem especially for certain age groups, with further variation by race and sex. In fact the census is used to adjust the birth certificate data that go into DA before DA is used to evaluate the census. Wachter also notes the complexity of DA,[21] and the fact that it is rightly subject to continual revision. He is particularly uncertain about the correctness of the estimates of immigration. He applauds the innovations in the 1990 DA, but quotes his colleague, Wolter, as saying: "The corrections that have been made are indicative of the corrections yet to be named."

The Post-Enumeration Survey and Dual-System Estimates

The Post-Enumeration Survey serves two related purposes. It is used as a measure of the accuracy of the census and it is used together with the census and statistical methods to generate adjusted counts. These adjusted counts are technically referred to as dual-system estimates (DSE). To evaluate the quality of the PES a series of 21 studies was done.[22] There are two questions that the Census Bureau intended to answer with the evaluation of the quality of the PES. First, whether the survey itself was of high enough quality in design and operation to be able to tell us something reliably about the faults of the census. Second, whether the adjusted counts or DSE were significantly more accurate than the census.

The Quality of the PES Survey

The 21 Census Bureau studies were designed to address the issues of quality in the PES and the DSE, some in a quantitative way, some in a qualitative way. They generated volumes of data that have not yet been fully analyzed or understood. Nevertheless, they have generated the basic material on which a judgment must be made regarding a possible adjustment of the census and the effect of that adjustment on the accuracy of the census.[23] In addition some of the panelists did their own studies on various aspects of PES quality. The broad picture that emerges from the analysis of these studies is that the PES was a generally high-quality survey that was well-executed.[24] There is little doubt that the PES detected an overall undercount in the census and a differential undercount at the national level by race and ethnic origin. But there are some problematic areas and disagreements among experts inside and outside the Census Bureau that have an impact on assessing the quality of the adjusted counts generated from the PES.

[17] Ericksen, page 2; Estrada, page 2; See also Appendix 3: Tukey, John W. "Recommendation on 1990 Census Adjustment," Member, Special Advisory Panel, June 18, 1991, (hereafter Tukey), page 3; Ericksen, *et al.*, pages 4–9.

[18] Nevertheless, this was at least the second-best census ever conducted.

[19] See the executive summary of these evaluation projects in appendix 2.

[20] See Appendix 13: Freedman, David A. "Adjusting the 1990 Census," Science, Volume 252, May 31, 1991, [hereafter Freedman] pp. 1233–1236.

[21] Wachter, pages 14–16.

[22] See the executive summary of these evaluation projects in appendix 2.

[23] Under Guideline Six, as explained later, "[i]f sufficient data and analysis of the data are not available in time to publish adjusted counts by July 15, 1991, a determination will be made not to adjust the 1990 census."

[24] See for example Ericksen, *et al.*, pages 12–16 for a good summary of the merits of the PES as a survey. Also see Wachter, page 2.

Missing data. The PES generates its estimate of the undercount by trying to match households it has information about to households in the census. A household survey in the PES that is matched to the census record of that residence means there was no undercount of that household. A non-match means there was an undercount. Matching is a difficult process and sometimes it is unclear whether there is a match or not. It is not an automatic process, rather it requires judgment and discretion. (For example, is a household headed by R. Smith the same as one headed by Bob Smythe?) Ideally, each household in the PES is matched or adequately resolved as not matched and thus missed in the census. Any case which is not resolved becomes "missing data" and, thus, whether those cases would add to or subtract from the undercount is unknown. The lower the missing data rate is, the more accurate the results are presumed to be. Three evaluation projects examined this problem.[25] In general, missing data were not found to be a serious problem; [26] however, there were two troubling findings. First, it is standard practice to impute persons into unresolved match households. The imputation rates for the two parts of the PES, called the "P" and "E" samples, were high: 1.7% and 2.1% respectively, which is equivalent to 3,900,000 and 5,025,000 individuals in the census when weighted up to the national population total estimate. These numbers are the same order of magnitude as the undercounts. Second, the percent of imputation in an evaluation stratum is highly correlated with the size of the undercount in that stratum. Thus, the strata for which there is more doubt about the quality of the adjusted data because of imputation tend to be the same strata for which an adjustment would result in large increases in the population.

Although Ericksen, Estrada, Tukey and Wolter do not find missing data or imputation to be a problem, Wachter raises some basic questions about imputation.[27] The imputation scheme used for the PES is based on a series of assumptions that are mostly guesswork.

Given the assumptions, Wachter finds that this work is of high quality, yet he is hesitant to believe that these assumptions are necessarily valid. To get some idea of whether the assumptions are important he calculates strict upper and lower bounds on the effects of imputation.[28] This analysis shows that if the imputation assumptions were incorrect, the variation in the estimates could be well beyond that expected from sampling error alone. Thus these untested assumptions are critical. They may in fact be correct, but if they are not, the adjusted estimates may be significantly in error. This implies that the estimates in the adjusted counts are subject to more potential error than has been computed.[29]

Matching error. Highly accurate matching is important because matching errors in even a small percentage of cases can significantly affect undercount estimates.[30] Ericksen, Estrada, Tukey and Wolter find the matching process to be of high quality.[31] Although Wachter does not dispute that this is what the studies show, he believes that the estimate for the matching error is too low, because the rematch study "does not, by its nature, expose certain inevitable kinds of matching errors." [32] For example, he notes that the structured nature of the PES interviews could lead to inaccurate and inflated estimates for undercount rates. His evidence, though anecdotal, is suggestive of the fact that the variance due to matching error is conservatively estimated in the total error model.

Erroneous Enumerations in the Census. Erroneous enumerations include people who died before or were born after census day, fictitious people and pets listed as members of a household, twice counted people as well as people enumerated outside the PES matching area. There were a large number of erroneous enumerations in this census and they were differentially distributed. "While the national rate of erroneous enumerations was estimated at 5.4 percent, the rate for Blacks, Hispanics and central city Asians was 7.7 percent compared to 4.4 percent for all others. Minorities in central cities had the highest erroneous enumeration rate at 8.4 percent." [33] The Census Bureau studies indicate that the PES was good at detecting erroneous enumerations, although three processing offices show statistically significant underestimates of erroneous enumerations.[34] The national effect of these errors is small, but the impact on regional totals is unknown.

Ericksen, Estrada, Tukey, and Wolter take the large number of erroneous enumerations as an indictment of the census.[35] Although it is certainly a matter of concern, especially for future census planning, the relevant question is whether the large numbers of erroneous enumerations would affect the accuracy of the proposed adjustment. Wachter considers this question at length.[36]

> Erroneous enumerations and cases with insufficient information are not part of the usual statistical framework for dual-system estimation. Their modeling has received much less attention than the omission rates . . . The PES, however, turns out to show that erroneous enumerations account for a large portion of the variations in net undercounts across areas and post-strata. *This outcome very much complicates the task of understanding and assessing the adjustment process.*[37] [emphasis in the original]

The adjustment factor for a post-stratum is determined by the netting out of two kinds of errors in the census—in technical terms, gross omissions minus erroneous enumerations. One would hope that the predominant determinant of the adjustment would be the number of people missed in the census: areas with high miss rates get high adjustments. What Wachter demonstrates is that the erroneous enumerations—the number of extra people counted—are what is really driving the adjustment: areas with low duplication rates get high adjustments. For example, the three regions with the highest omission rates have very different adjustment rates. Like Wachter, I find it disturbing that "erroneous enumerations account for a large portion of the variations in net undercounts across areas and post-strata." [38]

As Wachter notes, Ericksen, Estrada, Tukey, and Wolter take the high levels of erroneous enumerations as evidence that coverage improvement programs were not finding real people but just adding fictional people to the count.[39]

---

[25] See executive summaries of P1, P2, and P3 in Appendix 2.

[26] Estrada, pages 11–13.

[27] Wachter, pages 21–22.

[28] Wachter, page 22.

[29] In a letter submitted on July 11, 1991, Ericksen and Tukey dispute Wachter's concerns over imputation. Professor Wachter was offered an opportunity to respond in the interest of fair play. In his rebuttal letter, submitted on July 12, 1991, Wachter stands by his statements. Both letters are contained in Appendix 16. Wachter correctly notes that his claim was only that "a great deal rests on the correctness of the assumptions in the imputation," not that his alternatives were more reasonable than the ones used.

[30] Comments by Barbara Bailar. Journal of American Statistical Association. (March 19, 1985). Pages 109–111.

[31] Ericksen, page 13.

[32] Wachter, page 20.

[33] Ericksen, *et al.*, page 8.

[34] Estrada, pages 16–17; and the executive summaries of the evaluation studies P9 and P9a in Appendix 2.

[35] Ericksen, *et al.*, pages 7–9.

[36] Wachter, pages 12–14.

[37] Wachter, page 11.

[38] Wachter, page 11.

[39] Ericksen, *et al.*, pages 5–9 and Wachter page 12.

Wachter finds very mixed evidence on this question in comparing the counts in Detroit and Chicago. Late in the census enumeration, Detroit mounted an intense campaign to improve coverage, exceeding that mounted in Chicago. In the aggregate, Detroit did have a slightly higher erroneous enumeration rate, but a much lower omission rate. Thus, coverage improvement may very well have worked. However, for some categories of people, omission rates are roughly the same between the two cities, whereas erroneous enumeration rates are not. Thus, the evidence about coverage improvement is certainly more mixed than Ericksen, Estrada, Tukey and Wolter claim.[40] It is worth noting that Detroit and Chicago are lumped together when adjustment factors are calculated, despite their sizable differences in coverage patterns.

Correlation Bias. To the extent that the PES misses the same people that the census misses it will underestimate the undercount. The technical term for this problem is correlation bias. There are several ways of assessing the extent of this problem, but the basic message given by all of them is the same. There is strong correlation bias in the PES, especially among black males.[41] Ericksen, Estrada, Tukey and Wolter tend to dismiss this problem by noting that the presence of correlation bias results in an underestimate of the undercount, so an adjustment at least goes part way toward solving the problem.[42]

However, the presence of large correlation bias poses a fundamental difficulty for the adjustment procedure. Since there is no way to observe these people directly, the adjustment estimator attempts to include an estimate of these people. They are often referred to as the "4th cell" since they appear in the 4th cell of a 2 by 2 table in which persons in a particular post-stratum are classified as being in or not in the census and in or not in the PES. Unfortunately we have no direct data to verify if the assumptions for estimating the 4th cell are met. One piece of data indicates there may be a problem we do not fully understand. Traditional wisdom has it that males are generally more subject to correlation bias, since past data support the observation that males are more likely to be missed in both the census and the PES.[43] But, in 1990, about one-half of the people added to the estimate of the population from the 4th cell are women. Thus there is reason to doubt that the "fourth cell" numbers are correct. If that were the case the accuracy of the adjustment would be indirect.

One also expects that the number of people added to the adjusted population from the 4th cell should be small and that the estimate of the total population should be "lower than the truth." This is because no one expects that the estimate to fully reflect people missed in both the census and the PES. In past censuses, that has been the case. However, for 1990, the data are not consistent with past experience. Almost 5 million people were added to the estimate of the total population from the 4th cell, and the PES estimate of the total population exceeded the estimate from DA—a very unexpected finding.[44] Taken together, these findings indicate there may be problems in the adjusted count estimates that are not fully understood.

Wachter devotes several pages to the issue of correlation bias or as he calls it "catchability error."[45] His technical worry is that the allocation of this error to the model that measures the total error in the PES is done in an arbitrary fashion. Specifically, the national totals for black and for non-black males in six age groups estimated from DA are divided by the corresponding totals for females. Under the assumption of no correlation bias for females, these ratios are then multiplied by the national totals from the adjustment estimate for females in each group to give the predicted total for males. The differences between these predicted totals and the totals for men given by the calculated adjustment are the resulting national estimates of unreached persons. The method assumes all unreached people are men. This allocation, which critically affects conclusions about the accuracy of the census, is not based on empirical evidence on the distribution of those persons not reached by either the census or the PES, but rather on a formula of convenience. There is no unique way of choosing an allocation scheme. The one chosen is not obviously bad, but whether it is good is speculative and has no basis in fact. Furthermore, the variation in the PES estimates contributed by correlation bias is computed for sex ratios in an "ingenious" but ultimately untenable fashion.[46] It uses the capture probability of those reachable by the PES and census to infer a capture probability for people who intend to evade both the census and the PES.[47]

Wachter's argument over this technical point takes him back to a more fundamental point raised earlier, and also raised by Special Advisory Panel Members Kruskal and McGehee.[48] The PES is based on a statistical technique called "capture-recapture" which is often applied to estimating wildlife, particularly the number of fish in a pond. Fish are caught, tagged, thrown back and some are recaught in a second catch. An estimate of the population of fish can be made from the number of fish who are tagged on the second catch. The analogy made for the adjustment mechanism is that the census is the first catch and the PES the second. The analogy is not close, and it is not routine to adapt the wildlife model to counting the population.[49] The problem that

[40] Wachter, pages 12–13.

[41] See the discussion above.

[42] For example, see Estrada, page 14.

[43] See the discussion of hard to count groups in C.E. Citro and M. L. Cohen, eds., The Bicentennial Census, National Academy Press, 1985, Chapter 5, especially pages 177–186, and pages 224–237.

[44] See also the earlier discussion regarding the differences between DA and the PES at the national level.

[45] Wachter, page 18.

[46] Wachter, pages 18–19.

[47] In their letter submitted on July 11, 1991, Ericksen and Tukey dispute Wachter's concerns over the consistency of DA and the PES. In his rebuttal letter, submitted on July 12, 1991, Wachter stands by his statements. Both letters are contained in Appendix 16. It is difficult to referee this dispute at the eleventh hour, especially since the lateness of the Ericksen/Tukey letter gave little chance for Wachter to prepare a detailed response. It seems however, that even given the recognized inability of the PES to reach certain black males, a PES-based adjustment would have more persons than demographic analysis would indicate. Now suppose, in fact, that one were to use the behavior of those captured by the PES to extrapolate to those missed by both surveys, as Ericksen and Tukey suggest. The estimate of the population would be, at least by Wachter's estimate, yet another half-a-million higher. Then the PES would exceed DA by well over a million people.

Ericksen and Tukey also take Wachter to task for asserting that "[t]here is no evidence we know of that indicates that a substantial proportion of those persons counted neither by the PES nor the census avoided being counted." Ericksen and Tukey have apparently overlooked a well known study by Valentine and Valentine that concludes "one cannot [always] expect traditional interview or self-enumeration procedures to identify individuals of the type missed in the study area. * * * [T]he men were not reported because identification * * * could be detrimental to the economic welfare of the household." Citro and Cohen, *op cit.* pages 236–37.

[48] See Appendix 3: Kruskal, William, "Recommendation to the Secretary on the Issue of Adjusting the 1990 Census," Member, Special Advisory Panel, June 13, 1991, [hereafter Kruskal], page 2; Wachter pages 18–20; and also Appendix 3, McGehee, J. Michael, "Report to Secretary Robert A. Mosbacher on the Issue of Adjusting the 1990 Census," Member, Special Advisory Panel, June 21, 1991, [hereafter McGehee], pages 8–12.

[49] Citro and Cohen, *op cit.*, page 147, make this point clearly.

worries both Wachter and Kruskal is that, using the fishing analogy, some fish are harder to net than others.[50] There are, among fish, some "wily trout" which cannot be caught at all. Similarly some persons are harder to count than others, and some impossible.[51] For a variety of reasons they avoid the census and other forms of registration. The conclusions drawn about the population depend on what assumptions are made about these unreachable people. Different assumptions lead to widely differing results.

McGehee's concern about the application of capture-recapture is related to this notion of countability. The census and enumeration are both done by enumerators of varying skills, in different kinds of geographical areas (urban, rural, inner city, suburb) in an attempt to enumerate people who have different incentives to cooperate with the census or the PES. Thus there is inherent in the process a large variation in the probability of a particular person being enumerated in a particular place by a particular census worker. Further, to see if a person was counted both in the census and the PES a match has to be made—we do not tag people like we tag trout.[52] The ability of the matcher thus comes into play here. McGehee recognizes that there are elaborate mechanisms in place to control for all the potential variation, but many of those mechanisms depend on unverified statistical assumptions about what is important, and are changed after the data are in or after new research is completed.[53]

Total Error Model. An effort was made to produce estimates of expected error in the PES and variability of the estimates derived from the PES in project P16. This is generally referred to as the *total error model* since it was an attempt to combine the errors found in the PES by the other evaluation studies. These estimates of error cannot be made for any detailed groups. Instead, the population is divided into thirteen very broad categories called evaluation strata.[54] The estimates of errors for each evaluation strata are meant to be indicative of the uncertainties due to sampling error and all known components of non-sampling error. Whether the results of this study of large groups holds for smaller groups such as post-strata, states, cities or districts is unclear.[55]

This evaluation technique represents pioneering work on the part of the Census Bureau. It has been refined several times since the beginning of June, and every indication is that more refinements will be made as research on it is completed over the next several months. Nonetheless, some conclusions can be drawn from this project. On the one hand, the errors introduced by measured flaws in the PES process seem small. On the other hand, the model does show that the PES is biased toward overestimating the undercount and that a bias-corrected estimate of the undercount would be about 1.4 percent rather than the production estimate of 2.1 percent. This means about a third of the net undercount adjustment in the DSE comes from bias in the PES.

Furthermore, the undercounts tend to be higher in the minority evaluation strata, as are the biases in the PES. Even after bias correction, the minority evaluation strata show statistically significant undercounts. Ericksen, Estrada, Tukey, and Wolter note that the shift in shares of each evaluation post-strata would be small if the production estimate were corrected for bias.[56] Wachter [57] expresses various concerns about the computation of the total error model and its components as does the minority of the Undercount Steering Committee.[58] The results of this model are used further in assessing the quality of the counts themselves.

The Quality of the Adjusted Counts

The fact that the PES was generally a high quality survey does not necessarily imply that it results in high quality adjusted counts. To the contrary, erroneous enumerations and correlation bias lead to the conclusion that there are serious doubts about the quality of the adjusted population estimates.

To understand the statistical issues involved in assessing the quality of the adjusted counts it is necessary to begin with a summary understanding of three measures of the population that the Census Bureau compared.[59] First there is the census enumeration. Second there are the adjusted counts or the production dual-system estimates (production DSE). Third there is an alternative DSE that corrects for biases found in the production DSE by examination of the evaluation of the PES in the P-studies. The third measure is used to judge the relative accuracy of the census and the production DSE. There are two main elements of concern: (1) whether to test the accuracy of population totals or of population distributions and (2) how such tests should be performed.

Should population totals or population distributions be compared? Acceptance of the PES measure of the national undercount as reasonable is only a necessary—not a sufficient—condition for it to be an adequate instrument to be used to adjust the actual enumerations. There has always been an undercount in the census. The central questions for the Constitutional and statutory purposes of the census are whether the undercount is evenly or differentially distributed across geographical areas and jurisdictions, and whether we know how to reduce the range of any differential undercounts. Indeed Congress has recognized this problem as well.[60]

These questions have not been squarely faced. For the most part, Census Bureau analysts concentrated on whether we know enough to reduce the errors in the numeric counts without regard to whether this increases or decreases the severity of differential undercounts across geographical areas or jurisdictions. That is, they interpreted accuracy as concerned with getting the number of people closer to the truth rather than getting the allocation of the population for the purposes of political representation and funding closer to the truth. The two do not necessarily go together.

An illustration of the problem with using the absolute criterion alone is useful. Suppose you observed an enumeration which missed exactly 5 percent of the people in each and every block. Thus, although 5 percent is missed in each and every block, the proportion of the total population in each block is still estimated correctly. Suppose now that you adjusted this enumeration by increasing the counts in half the blocks by 1 percent and increasing the counts in the other half by 5 percent. On average you would have reduced the undercount of the

[50] Kruskal, page 3; and Wachter, page 18.

[51] See Citro and Cohen, *op cit.*, pages 139–142.

[52] Although often trout lose their tags which poses a similar conceptual problem.

[53] McGehee, pages 8–12.

[54] A list of evaluation strata and their component post-strata are included in the Decennial Census Procedural Documentation, below.

[55] Wachter, page 16.

[56] Ericksen, *et al.*, page 15.

[57] Wachter, page 17.

[58] Undercount Steering Committee, page 6.

[59] These measures will be explained more fully in the course of this discussion. The alternative DSE is also called the "target" population in Census Bureau documents.

[60] Subcommittee Chairman Thomas Sawyer, for example, noted that "If the undercount were evenly distributed geographically and demographically across the population, it probably would not pose the problem that we confront here and the difficulty that we face in asking the Secretary to come to this decision." Hearing before the Subcommittee on Census and Population of the Committee on Post Office and Civil Service, U.S. House of Representatives, January 30, 1990. Serial no. 101–43. page 18.

population by 3 percentage points thus, improving the numeric accuracy of the nationwide total. The numeric accuracy of the absolute level of the count also would have improved for each block. However, the block proportions would now be wrong. Half the blocks would be 2 percent too small and half would be 2 percent too large relative to the average undercount. The absolute criterion would prefer this type of adjustment even though it moves from a situation in which every citizen gets his or her fair share of representation and funding to one in which every citizen got 2 percent too little or 2 percent too much.[61]

It is quite possible this kind of error could occur when the PES misses persons. The PES failure to include large numbers of black males in the adjusted counts could have caused just this kind of error. We simply do not know if it did.

I conclude that the Constitutional and legal purposes for the census must take precedence, and accuracy should be defined predominately in terms of getting the proportional distribution of the population right among geographical and political units. This argues for putting aside the judgment of accuracy based on getting absolute numbers right (numeric accuracy) and instead focusing on the question of whether there is convincing evidence that the accuracy of the population distribution in the adjusted numbers (distributive accuracy) is superior to the distributive accuracy of the actual enumeration. The quality of the adjusted counts themselves must be examined to address this important issue squarely.

What is the criterion for accuracy? Guideline One mandates that the census enumeration "shall be considered the most accurate count of the population of the United States, at the national, State, and local level, unless an adjustment is shown to be more accurate." This guideline requires a series of statistical hypothesis tests at various levels of geography in which the adjusted counts are to be presumed less accurate measures of the population than the actual census enumeration unless there is convincing evidence that the adjusted counts are closer to the true counts than the actual enumeration.

The true population counts cannot be observed. However, classical statistics provides a standard way of approaching the required inference. In accordance with Guideline One, we take as a working (null) hypothesis that the actual enumerations in fact better characterize the true population. The adjusted counts are an alternative measure and the question is whether the available evidence permits us to reject the hypothesis that the census better describes the true population.

We shall see below that the Census Bureau has provided substantial (although not necessarily "convincing") evidence that the adjusted counts are more accurate if accuracy is interpreted to mean numeric accuracy. However, the evidence provided by the Census Bureau tends to support the superior distributive accuracy of the actual enumeration. Thus, since accuracy is interpreted in terms of the fairness of the implied distribution of representation and funds, the Census Bureau report supports the conclusion that the adjusted counts are not more accurate.

The choice of accuracy criterion is crucial because there appears to be a substantial national net undercount in the numeric census counts. Simply correcting for the estimated net undercount can improve numeric accuracy but significantly worsen distributive accuracy. We can see that we missed people in most areas, but we lack a tool which can improve the distribution of population for the purposes of political representation and funding.

How are the tests of accuracy performed?

(a) The Census Bureau Loss Functions

The Census Bureau approach to testing the quality of the adjusted counts relies heavily on showing that the PES was well-executed and that the identified biases in the production Dual System Estimates or adjusted counts (DSE) are small relative to an "ideal" DSE. Unfortunately this type of validation methodology does not work in the present instance because of a basic design flaw: The DSE fits broadly into the class of "certainty-equivalent" predictors which use estimates as if they were known for certain rather than subject to statistical variation. A statistically optimal estimate of the population for an area would take account of this uncertainty.[62] Thus the conclusion that the measured shortcomings of the adjusted counts under consideration (the "production DSE") are small relative to the ideal DSE merely means that the production DSE has a chance of improving accuracy. It is unacceptable to go the next step and conclude that a good production DSE would be more accurate than the actual enumeration.

The production DSE are in fact less accurate than those ideal DSE because (a) the data were less than perfect, and (b) the correct model was not known. The bulk of the Census Bureau effort was aimed at seeing whether these data and modelling problems were disqualifying for the production DSE. It is clear that the production DSE are not unbiased estimates of the differential undercount rates and the DSE procedure overcorrects for the measured undercounts. This is measured in the total error model discussed above. These biases are quantified for thirteen large evaluation strata.

Using the total error quantification, the Census Bureau has generated an alternative Dual System Estimator of the population. It is worth noting here, that the errors in the production DSE are quantified for 13 very large groups of people. These errors are then "parcelled out" to the 1392 post-strata used to calculate an adjustment, the adjustment factors are corrected for these biases, and the alternative DSE is calculated. Since there are also estimates of variance for the DSE, the Bureau actually calculates a statistical distribution of possible alternative DSE. A thousand random draws from this alternative distribution were used to generate estimates which the Census documentation terms "the target population." This is not the true population distribution—which is unobservable—but rather a tool for assessing the quality of production DSE counts relative to an "ideal" DSE based

[61] Kruskal gives a similar example on page 7 of his recommendation. See also Citro and Cohen, *op cit.*, page 318. "While synthetic estimation is suggested for adjustment, because of its arithmetic and computational simplicity, synthetic estimation is not necessarily an improvement over the census count." Cohen and Citro use a numerical example as an illustration.

[62] The optimal estimate would average the ideal DSE estimate (based on the correct model and perfect data) with the actual enumeration with more weight being put on the actual enumeration when the model parameters are less precisely estimated. In point of fact, there are statistical theorems which demonstrate that even if the correct statistical model were known and perfect data were used, the Dual System Estimator (DSE) could generate adjusted counts which are either (1) clearly less accurate, or (2) not significantly more nor less accurate, or (3) clearly more accurate measures of the true population than the actual enumeration from the census. The question is which of these occurred? A textbook analysis of the suboptimality of the certainty equivalent approach is found in Arnold Zellner, An Introduction to Bayesian Inference in Econometrics, New York: John Wiley & Sons, 1917, pages 322–327. Intuitively, the problem arises because a full correction is attempted which is optimal only if one knows exactly the undercount or overcount in each area. As the actual uncertainty increases about exactly where and how many people were missed, attempts to make the full estimated correction increase the error variance relative to optimal and eventually, if uncertainty is large, relative to the unadjusted counts.

on more perfectly measured data and more correct models. But this hypothetical DSE is also just an estimator—subject to statistical error. So a correct analysis must account for two errors: (1) the error that comes from using the production DSE rather than the idealized DSE and (2) the error that is inherent in the idealized DSE. Then that combined error should be compared with the error in the actual enumeration.

To make matters even more complicated, legislative—and, now, judicial—representation must be apportioned and allocated over many levels of government into districts that treat their residents as fairly as practicable. Thus, comparisons must be made not only at the various levels of government on which funding is based, but down to the census blocks which are the basis for drawing district lines for Federal, State, and local elections. Unfortunately, the Census Bureau did not have the time to conduct the hypothesis tests required by Guideline One before the Undercount Steering Committee report was completed on June 21, 1991. The method they used instead to make these comparisons is called loss function analysis.

In brief, loss function analysis is used to compare two sets of counts for the same population. Ideally, one of the sets of counts is the true population, and thus the loss in accuracy from using the alternative set of counts is measured. In practice, however, the truth is not known, so care must be taken in the interpretation of results. A loss function analysis can be performed at any level of geography—states, counties, cities, precincts, or blocks.

As an example, suppose a loss function analysis is being calculated for states. The difference between the two estimates of the population is calculated for each state. Then some kind of average is taken of the differences across all states to get an aggregate measure of total loss. The differences may be squared, summed and the total divided by the number of states. Alternatively, the absolute values of the differences may be averaged, where the average is weighted by the size of the state. There are an infinite number of formulas that can be used to average the state-by-state losses to get a single measure of total loss. These formulas are called "loss functions," and the results of any analysis can depend heavily on which loss function is chosen. For example, the loss function that uses squared differences penalizes a few large errors much more heavily than many small errors. The absolute value loss function does not have this property. The choice of a loss function is not scientific. It is usually made on the basis of convenience or tradition.

One more general comment on loss function analysis is needed. The loss function is ideally suited to measuring loss when an estimator of a population count is being compared to a known true count. In this case, the interpretation of the loss is straightforward. It is the accuracy lost by using the estimator. However, when one imperfect estimator is being compared to another, it is more difficult to interpret the loss of one estimate. The temptation is to call one estimator the "truth" and measure loss against it. But one is not measuring loss against the truth. This is simply measuring loss of one estimate against another. There is no reason to think this analysis tells you anything about the truth. In loss function analysis, it is critical to consider the base being used for comparison—losses are measured only relative to that base.

The loss function analysis run by the Census Bureau asked whether the enumeration or the production DSE was closer to the "ideal" DSE.[63] This does not form a statistical test of whether the production DSE are more or less accurate than the census counts. It only calculates which set of numbers on average is closer to another set of estimates (the target population). These tests were simply not proper statistical tests to address the critical hypothesis about the distributive accuracy of the PES and the census enumeration.

Their examination of this closeness question erred further in two significant ways: (1) Instead of comparing the production DSE that would be used, they compared the mean of 1000 draws from a model reflecting the statistical properties of the DSE. This effectively eliminates the inaccuracies derived from using one particular set of adjustments. (2) Rather than using Guideline One's mandate that the actual enumeration be deemed more accurate unless the adjusted counts are shown convincingly to be more accurate, the Census Bureau did the reverse—they preferred adjusted counts if the actual enumeration was not proven more accurate.[64] Thus the Census Bureau loss function analysis was seriously deficient.

There is, nonetheless, a June 27, 1991, Addendum to the Undercount Steering Committee report of June 21, 1991, that corrects some initial flaws in the loss function analysis.[65] This addendum attempts to correct for the error in failing to allow for the fact that the target population was itself an estimator subject to random variance. An allowance for this variance was removed from the variance charged to the census counts and estimates made of the number of states for which the population proportion would be made less accurate was generated. The number of state proportions worsened depends crucially upon the allowance made for variance in the alternative DSE: If only the variance measured in the total error model is used, then the shares of an estimated 21 states are made worse by adjustment (using an absolute value loss function).[66] However, this is clearly a minimum estimate. "As a matter of judgment, the total understatement of variance of the estimates from the smoothing model may be in the range of a factor of 1.7 to 3.0 in terms of variance," according to the Undercount Steering Committee.[67] Allowing for a variance factor of 2.0, which is near the lower end of the Undercount Steering Committee range, the proportional shares of about 28 or 29 states would be worsened by an adjustment in terms of distributive accuracy.[68]

Even with the variance factor set at only 1.0, adjustment is estimated to have worsened distributive accuracy compared to the census counts in 11 of the 23 metropolitan areas in cities with 500,000 or more persons: Phoenix, Washington, DC, Jacksonville, Chicago, Baltimore, New York City, Memphis, Dallas, El Paso, Houston, and San Antonio. Again using only the measured variance, half of the 14 metro areas in counties with over 500,000 persons are made less accurate proportionally by

---

[63] This loss function analysis is described in detail in Undercount Steering Committee, pages 6–7; and Bryant, pages 12–14.

[64] This last error may reflect the fact that the Census Bureau ignored the difference between the true population and its own approximate ideal estimator. See for example, the Undercount Steering Committee, page 2: "Time did not allow for full simulations of accuracy for smaller areas. There was some evidence from the loss function analysis, but there was no independent evidence with which to compare it. . . . Even so, in the absence of direct evidence to the contrary, the majority concludes that adjusted counts are generally more accurate at lower levels."

[65] A discussion of how this change affected the Undercount Steering Committee's conclusions is contained in the discussion of Guideline Six, below

[66] See Appendix 5. Addendum to the Undercount Steering Committee Report, July, 1991. [hereafter Addendum], page 3. Given the original erroneous analysis, the Undercount Steering Committee report (page 6) was formulated when the committee thought the accuracy of only about 11 states was worsened by adjustment.

[67] Undercount Steering Committee, page 5. The actual variance is believed to substantially exceed the measured variance because of doubts similar to those raised by Wachter for the matching and imputation procedures.

[68] Addendum, page 4.

adjustment. Only aggregate measures are available for areas of other sizes. These show that on average the adjusted figures improve distributive accuracy relative to the census, but no detail is given as to the number of jurisdictions for which the PES is closer than the census. In all these sub-state cases, too, the estimated distributive accuracy of the adjusted figures deteriorates dramatically compared to the census if the variance is increased to allow for the unmeasured uncertainty in the estimator.

In sum, the corrected Census Bureau estimates of distributive accuracy marginally favor the adjusted counts—though many states and communities would be less accurate—*if* only the measured variance is used. When the variance is increased into the plausible range (in the professional judgment of the Undercount Steering Committee), distributive accuracy comparisons are more favorable to the census counts.

It is worth reiterating that Guideline One specifically places the burden of proof on the adjusted estimates, not on the census. The census is considered to be more accurate unless the adjusted figures are shown to be more accurate. With respect to places under 100,000 population there is no direct evidence that adjusted counts are more accurate.[69]

What evidence there was based its conclusions primarily on the numeric accuracy of the adjusted counts rather than the adjusted proportions, and that the Bureau depended upon indirect evidence rather than direct tests of statistical hypotheses.[70]

(b) Face validity tests

In addition to Loss Function Analysis computed by statisticians, demographers at the Census Bureau made an independent evaluation of the adjusted population counts for states. To do this they compared the adjusted state counts with counts simulated by DA. To make the simulations (because DA provides data only at the national level), they disaggregated census counts for each state by race and Hispanic ethnicity. They then applied DA national undercount rates to black and non-black subpopulations and PES rates to Hispanic and Asian and Pacific Islanders. Then they built up new state estimates by recombining the racial and ethnic groups. These simulated state estimates further confirmed the "face validity," or reasonableness, of the adjusted state counts.[71] These face validity tests depend critically on what the analyst expects. Face validity tests certainly cannot be a substitute for formal tests, but just as face validity can be used to show that adjustment is making counts more accurate, face validity can show the opposite.

For example, is it reasonable that New Mexico has the highest undercount rate of any state? Why should the undercount rate for Montana be higher than that of New York State? How can the very low estimated undercount rates in cities like Philadelphia be explained? Of the large cities, only Washington, DC and Boston showed increases in their black populations between 1980 and 1990. Yet, Washington DC is estimated to have a very large undercount rate and Boston is estimated to have a very small undercount rate. Why are the only estimated overcounts for cities over 100,000 concentrated in New England? Why should Akron and Dayton have high estimated undercount rates (3.0% and 3.3%, respectively) and Cleveland have such a low estimated undercount rate (1.4%)? These examples illustrate as above the point noted above that was raised by Wachter earlier—there is much more texture to the pattern of undercount that lies well beneath the surface of any aggregate loss function analysis. Face validity cuts both ways. And the face validity of the proportions of persons in states and localities has not even been checked.

(c) Ericksen, Estrada, Tukey, and Wolter's claims regarding accuracy

These panelists take a different approach to the problem of accuracy of the counts at state and local levels. An article by Wolter and Causey attached to their jointly authored document [72] argues that accuracy improves, on average, at lower levels, so long as the measured undercounts at aggregate levels tend to have smaller errors than the original enumeration. In addition it is argued in a similar manner in an attachment to the joint report that Ericksen, Estrada, Tukey and Wolter submitted that adjusted counts will on average improve block level data (and thus data for localities) consistent with its improvement of data at larger units of geography.[73] Thus their argument asserts that by applying the total error model to the 13 evaluation post-strata, the PES is shown to be more accurate than the census and the error in the PES is shown to be low. They conclude, based on the theoretical argument by Tukey and the empirical argument made by Wolter and Causey, that

a. The total combined error increases as the size of the group decreases; *e.g.*, the combined errors for 5 million blocks will be larger than the combined errors for 1392 post-strata.

b. Consequently, the improvement in amount due to adjustment will be nearly the same for larger and smaller groups, the improvement in percentage terms decreases, but does not change sign, as the groups become smaller.[74]

Ericksen, Estrada, Tukey, and Wolter note that these conclusions depend on a particular measure of combined error—a loss function that uses a size-weighted sum of relative error. Their primary point is that, with such an error measure, conclusions about local accuracy can in fact be drawn from accuracy conclusions at larger levels. In short, they contend, "improvement in quite large areas thus prophesies improvement in very small areas, as well as a variety of intermediate levels." They see a post-enumeration survey with small measured error (and some, like Wachter and the Undercount Steering Committee contend that such error is very conservatively measured) for thirteen large evaluation strata. They conclude that the adjusted counts for these large evaluation strata are more accurate—a questionable inference because they made no formal statistical test of this hypothesis. From this

[69] The Undercount Steering Committee report states "in the absence of direct evidence to the contrary, the majority concludes that adjusted counts are generally more accurate at lower levels," and later "while analysis was not available for smaller areas, the majority concludes that acceptable patterns would happen there also." (Undercount Steering Committee, page 2.). The reasoning is contrary to the explicit mandate of the guideline. Similarly the Director stated, "there is little evidence to judge whether the proportional distribution of adjusted counts is more accurate for places under 100,000. However, Loss Function Analysis shows that for metropolitan places of less than 25,000, 25,000–49,000 and 50,000 or more, and for nonmetropolitan places less than 25,000, and 25,000–49,000 in total, by these sizes categories, adjusted counts are more accurate than the census. However, there are concerns about the accuracy of the loss function assumptions for small areas." (Bryant, page 14.)

[70] In a June 28, 1991, memorandum Senior Mathematical Statistician Robert Fay reports his efforts at conducting formal hypothesis tests of the distributive accuracy of the adjusted figures at the state level only. There was not time for the Undercount Steering Committee to review this memorandum and it may contain further errors. Nonetheless, although the hypothesis tests rejected the superior distributive accuracy of the census counts if only the measured variance was changed to the adjusted figures, the superior accuracy of the census counts was easily accepted for a variance factor of 2.0 and appears (by interpolation) acceptable at any variance factor in the Undercount Steering Committee's plausible range of 1.7 to 3.0.

[71] Bryant, page 14.

[72] See appendix G of Ericksen, *et al.*

[73] See appendix F of Ericksen, *et al.*

[74] Ericksen, *et al.*, page 20.

questionable conclusion they apply mathematical theory to infer *average* accuracy improvements at lower levels.

In testimony before Congress, an official of the General Accounting Office raises some questions on the issue of sampling error and lower level geographic accuracy:

> We believe the amount of sampling error, or variability, deserves attention by the Secretary because it was a consistently high source of uncertainty in the PES over- and undercount estimates. The PES estimates are based on samples and therefore subject to random error. The levels of sampling variation measured by the evaluations of the PES were generally much higher than anticipated by the original design of the PES. For example, even after smoothing to reduce sampling variability, PES over- and undercount estimates for 4 of the 13 evaluation groups did not show a statistically significant difference from the census count. In other words, due to the variability resulting from doing a sample, *the Secretary cannot be sure whether 4 of the 13 population groups reviewed in the Bureau's evaluation of total error in the PES were overcounted by the census, undercounted, or if the census count was correct.* (emphasis added)
>
> The need for precision is especially important because the Bureau's procedure for carrying down PES adjustment factors to lower geographic levels applies the same adjustment factors to large numbers of people over wide geographic areas with similar demographic characteristics.[75]

The Wolter/Causey paper does not address this argument directly. In addition, Wachter argues cogently against indiscriminate use of the Wolter/Causey paper:

> Theirs is a very interesting paper, but its relevance is limited by its concentration on highly aggregated summary measures of improvement. It does not present explicit results on how many units at various levels might be made worse and how many made better by an adjustment. Furthermore, important calculations in the paper depend on stylized assumptions about correlations in the components of the undercount which may or may not hold in fact either for previous PES-like data or for the 1990 PES. These prior studies are valuable, but they are no substitute for examination of the actual 1990 data.[76]

There are fundamental difficulties with the Wolter/Causey argument. I am not convinced that at the evaluation strata level we can conclude the PES is more accurate. First, the measured bias alone is one-third of the total undercount and the Undercount Steering Committee itself stated that there are other non-measured sources of error.[77] Wachter also raises several fundamental concerns about this measurement. Second, the analysis depends on a particular loss function that weights a few large relative errors more than many small ones. This is not inherently bad, just arbitrary.[78] Wachter perhaps summarizes it best:

> *I do not believe that any highly aggregated index or loss function is appropriate for summing up overall accuracy.* It is informative to understand how much the outcomes of calculations with different versions of such aggregated indices differ. But the choice among them is not a scientific choice. Each such index involves implicit value judgments about different sorts of error. For example, each index determines whether a few large errors are more serious that a great many smaller errors. *Whether we agree with a particular tradeoff is a matter of personal and political values. It should not be disguised as science.*[79] [emphasis in original]

Loss functions mask the incredible complexity of the adjustment operation behind a single number. To get a glimpse of this complexity, it is useful to look at the undercount rates by state. Table 1 and Figure 2 of Appendix 10 show the undercount rate by state with margins of error. Counting the District of Columbia as a state, 42 of the 51 states show an undercount rate that is statistically significant. More importantly, however, is how these undercount rates differ from the national average, since it is these differences that determine which states win and which lose. Table 6 and Figure 1 of Appendix 10 show these differences again with margins of error. Only 18 of the 51 states have an undercount rate that is significantly different from the national average. That means in 33 states we do not know if the undercount rate is higher, lower or the same as the national average. Put another way, we do not know if these 33 states deserve more or less political representation and Federal funding than they are receiving. We do not know for these 33 states if an adjustment would result in a more equitable distribution of political representation and resources.

There are winners and losers from an adjustment—that is to be expected whenever a fixed set of resources is going to be divided. More seriously, however, there is general agreement that there will be some localities' counts that will be made less accurate by an adjustment. The proponents contend that, on average, more areas are made accurate, or more people live in areas whose counts are more accurate, or on average the counts are more accurate. These are all vague and general statements that do not describe the areas of the country where accuracy is likely increased and decreased, the types of towns where accuracy is likely increased and decreased, the neighborhoods where accuracy is likely increased and decreased. We have already seen above that general statements about improved accuracy on average are little if at all justified if realistic values are used for the error variance of the alternative DSE. Furthermore, the rhetoric, if not always the analysis, is centered around absolute levels of the counts, not improvements in the distribution of the counts.

Conclusions

Guideline One requires that convincing evidence be offered that the adjusted estimates of the population are more accurate than the census at the national, State, and local levels. In the absence of such evidence, the census counts are concluded to be the most accurate.

At the national level, it is likely that the PES-adjusted estimates reflect more accurately the total population and the racial and ethnic populations of the country. It appears equally clear, however, that the PES omitted large numbers of certain groups—notably black males. We have no information on the location of these persons. In addition, the PES and demographic analysis lead to sharply different conclusions about the accuracy of the census for several age/sex groups at the national level. Although these are not definitive disqualifiers at the national level, they do raise some question as to whether the adjusted figures are more accurate than the census count even at the national level.

The Constitution requires a census every 10 years not just to count the total number of people in the United States but to locate them so that political representation can be allocated to the states and the people in them in proportion to their numbers. I conclude that the primary criterion for accuracy should be distributive accuracy—that is, getting most nearly correct the proportions of people in different areas. Improved numeric accuracy, although in itself desirable, cannot compensate for treating states and individuals less fairly.

---

[75] See appendix 17. General Accounting Office. "1990 Census: Applying PES Results and Evaluations to the Adjustment Decision." Testimony before the Subcommittee on Census and Population, Committee on Post Office and Civil Service House of Representatives. [hereafter GAO Report]. Pages 7–8.

[76] Wachter, page 2.

[77] Undercount Steering Committee, page 5.

[78] Indeed, Ericksen, Estrada, Tukey, and Wolter make no claim of uniqueness for their choice of loss function. As noted earlier, the choice of loss function can control the results of an evaluation.

[79] Wachter, page 5.

At the State and local level the correct statistical analysis for both distributive and numeric accuracy simply has not been completed. The total error model indicates that the adjusted figures tend to be too high but generally closer in numeric terms to the true population than the census counts which tend to be too low. However, there is sufficient uncertainty about the true variance of the adjusted figures that even numeric accuracy has not been definitively demonstrated. The loss function analysis and hypothesis tests that have been prepared by the Census Bureau to date, although of uncertain reliability, do support the superior accuracy of the census counts versus the adjusted figures when we consider distributive accuracy—or fairness—and use reasonable estimates of the error variance of the alternative DSE. That is, for the Constitutional purposes of the census the available evidence is consistent with the census counts being more accurate than the adjusted counts. There is certainly not sufficient evidence to reject the distributive accuracy of the census counts in favor of the adjusted counts.

I conclude that, in accordance with Guideline One, the census counts are the most accurate count of the population of the United States at the State and local levels. While the preponderance of the evidence leads me to believe that the total population at the national level falls between the census counts and the adjusted figures, that conclusion is not relevant to the determination of distributive accuracy. Thus this guideline weighs in favor of a decision not to adjust.

Guideline Two

The 1990 Census may be adjusted if the adjusted counts are consistent and complete across all jurisdictional levels: national, State, local and census block. The resulting counts must be of sufficient quality and level of detail to be usable for Congressional reapportionment and legislative redistricting, and for all other purposes and at all levels for which census counts are published.

Explanation

This guideline acknowledges that the population counts must be usable for all purposes for which the Census Bureau publishes data. The guideline also reinforces the fact that there can be, for the population at all geographic levels at any one point in time, only *one* set of official government population figures.

Thus, the level of detail must be adequate to produce counts for all such purposes. If the 1990 Census count is to be adjusted, it must be adjusted down to the census block level. It must be arithmetically consistent to eliminate confusion, and to prevent any efforts to choose among alternative sets of numbers to suit a particular purpose.

If the Census is to be adjusted, a process called synthetic adjustment will be used. A synthetic adjustment assumes that the probability of being missed by the census is constant for each person within an age, race, Hispanic origin, sex, and tenure category in a geographical area. A synthetic adjustment is performed in two steps. First, the preferred adjustment factors are estimated for a variety of post strata defined by age, race, Hispanic origin, sex, and tenure within geographic areas. Then the adjusted estimate in each category for a census block is obtained by multiplying the unadjusted census estimate in that category by the adjustment factor. The adjusted census estimate for the census block is computed by adding the estimated adjustments for each post strata cell of the block. Put simply, in an adjusted population count each individual enumerated will receive a relative weight according to his or her race, age, sex, ethnic background, tenure, and place of residence. The aggregate counts will then be built up from the weighted individuals to census block, local area, state and national counts.

We will conduct evaluations of small area estimations to ensure that this process results in counts that are in fact more accurate.

Evaluations of small area estimation. Coverage error may vary substantially within the post-enumeration-survey post-strata, although the post-strata were drawn to be homogeneous with respect to expected coverage error. The goal of this analysis is to determine whether or not the assumptions underlying a synthetic adjustment of the census are valid and produce counts which are more accurate at all geographic levels at which census data are used. In particular, the within-strata block-to-block variance in characteristics and net overcounts or net undercounts will be analyzed.

Discussion

If I had determined that an adjustment should have been undertaken, the Census Bureau would have issued block-level Public Law 94–171 tapes that would have replaced those issued in the first three months of this year. Replacement Summary Tape File (STF) data would have also been issued and all future census products would have used adjusted counts. Our ability to have done so would have satisfied the production requirements of this guideline.

The substantive question here is whether the adjusted counts are of sufficient quality to be used for all purposes for which census counts are published. Clearly the quality of the adjusted figures is intimately related to their accuracy, which, as the discussion of the preceding Guideline shows, does not compare favorably with the actual enumeration. This Guideline raises another issue—synthetic adjustment.

As explained earlier, the adjustment process uses a survey of persons in 5,290 block clusters to change the number of people in 4,830,514 blocks. Based on extrapolation from this survey 6,188,204 unidentified persons are added by duplicating records of people counted in the census, and 918,937 people who were actually counted in the census are deleted. The adjustment process is done by dividing the population of the country into 1392 groups. Each member of one of these groups is assumed to have the same probability of being missed in the census as every other member of that group. The real quality of the census in a given block or even a given city has little impact on the adjustment of the count of the population of that block or that city. As will be seen in the discussions of Guidelines Seven and Eight, most local officials think that the adjustment will fix particular problems that they have identified in the count for their towns. It would do no such thing.

A synthetic adjustment assumes that the probability of being missed by the census is constant for each person within an age, race, Hispanic origin, sex, and tenure category in a geographic area. These groupings of persons are called post-strata. A synthetic adjustment is performed in two steps. First, the preferred adjustment factors are estimated for 1392 post-strata. Then the adjusted estimate in each category for a census block is obtained by multiplying the unadjusted census count in that category by the adjustment factor. The adjusted census estimate for the census block is calculated by adding together the estimated adjustments for each post-strata represented in the block. Because of the problems of correcting a census with a survey, adjusted figures cannot be more accurate than the census counts in each of the 4,830,514 occupied blocks, or at all larger aggregations of them. There is no PES system—short of one which took a second perfect census—that could say adjusted counts are more accurate for all blocks. The question is whether the assumptions that underlie this synthetic

adjustment mechanism are good enough to conclude that the counts are sufficiently accurate to be usable at a block or precinct level.

As noted above, the synthetic adjustment process rests on the assumption that persons in each post-stratum are homogeneous with respect to their probability of being missed by the census, *i.e.*, their capture probability. This is admittedly a very difficult thing to measure. There were several approaches taken by the Census Bureau to validate the homogeneity assumption, all contained in project P12.

The first part of P12 collapsed the 1392 post-strata by age and sex into 116 larger groups. To test whether the people living on blocks within these 116 larger post-strata are homogeneous with respect to capture probability, the Census Bureau conducted an analysis of the homogeneity of 115 of the 116 larger post-strata (the 116th is persons living on Indian reservations). A regression model predicted an adjustment factor for block parts, then compared that with an adjustment factor of 1.0 (no adjustment) representing the numeric census counts. This predicted adjustment factor was also compared with the measured factor for the post-strata used in creating the adjusted counts. For 24 of the 115 post-strata the census count was superior while for 91 post-strata the adjusted count was superior in terms of numeric accuracy. The Director interprets these findings to "give support to the accuracy of the selected PES adjustment model." [80] Regrettably, this evidence does not directly address the homogeneity issue. Like the uncorrected loss function studies this simply compares the census and the PES to yet a third estimate (the regression equation) whose quality or closeness to truth is unknown. This cannot be called a test or even a verification of the homogeneity assumption. To pursue this approach, allowance should have been made for the true variance in the regression estimates in a manner analogous to that done in the Undercount Steering Committee Addendum for the target population. It must be understood that such errors can easily occur when cutting edge research is used for production purposes under extreme time pressure.

The second part of P12 analyzed the homogeneity of state parts within post strata. Techniques known as analysis of variance were used to determine the validity of using post-strata, rather than states, for estimation of adjustment factors. The study was designed to determine if there was relatively more homogeneity within state or within post-strata. The study showed that, with the exception of the Mid-Atlantic Division, state differences were not significant within post-strata. This result was compatible with the conclusion that there is relative homogeneity for state parts within post-strata. There is no evidence of homogeneity for other geographic levels. [81] The only conclusion that can be drawn from this study is that the Census Bureau was better off using the actual post-strata for synthetic estimation than using any state-specific effects. Whether the levels of homogeneity within post-strata are acceptable is not even addressed.

The third part P12 looked at state homogeneity from a different vantage point. It measured whether other factors that are often correlated with undercounting are homogeneous within post-strata. Contrary to the results of the second part of the studies, these factors showed significant heterogeneity by state within post-stratum for well over 80% of the post-stratum groups. This study went further and measured the homogeneity of some of the components of the dual-system estimates at the block level and found about 14% of the post-strata groups to have significant state effects. Thus, the evidence in this study for the presence or absence of homogeneity within post-strata is mixed.

In summary, the analysis presented for decision from P12 was substantially different from that planned by the Census Bureau and used only the State as a surrogate for heterogeneity. We clearly do not thoroughly understand whether or not heterogeneity is a real problem. There are indications that using post-strata for synthetic adjustment is better than using states, but nothing more. It is impossible to conclude from any information the Bureau has presented in P12 that there is not residual heterogeneity within post-strata. [82]

Project P15 approaches the homogeneity problem by attempting to measure the quality of the dual system estimates by examining their expected variability. The measure used to do this is called the coefficient of variation which is the ratio of the sample standard deviation to the sample mean. The PES was designed so that these coefficients of variation were expected to be equal to 0.7 percent for the areas used in the design. In fact, in 48 of the 54 areas examined, the actual coefficients of variation are larger than expected. They ranged from 0.45 percent to 4.4 percent. This is direct evidence of substantially more variability in the DSE than expected and indirect evidence of heterogeneity within post-strata.

Other arguments have been made about this guideline. As noted in the analysis of Guideline One, Ericksen, Estrada, Tukey, and Wolter rely heavily on the Wolter/Causey/Tukey argument that synthetic adjustment will increase the accuracy of the counts. [83] For the reasons explained in the discussion of Guideline One, I do not find this argument compelling. Its reliance on the unsubstantiated homogeneity assumption simply emphasizes the concerns raised earlier.

Estrada argues that it is not necessary to show that the adjusted counts have to be better for all purposes, if it is shown on average to improve counts for its principal uses. "Improved counts to meet Constitutional needs for reapportionment and redistricting would be sufficient justification to adjust, even though for some other uses adjusted counts are less valid." [84] I do not consider this argument persuasive. Reapportionment and redistricting counts are the most demanding in terms of accuracy because block level counts are required to accomplish both. [85] If adjusted counts are better for these purposes, then they would necessarily be better for all others.

McGehee asserts that "variances between processing offices and evaluation strata fall outside expected levels in a number of the evaluation studies. At the district office level and below the data contain such wide variances that they could not be reconciled without weighting them to

[80] Bryant, page 18.

[81] The Undercount Steering Committee report states that a majority of the Undercount Steering Committee believe this result would hold for other geographic levels. However, there is no evidence presented to support this. Undercount Steering Committee, page 2.

[82] Estrada agrees that the findings from P12 are mixed, although his conclusions differ from mine: "It supports the fact that PES poststrata are homogenous with respect to expected coverage error, but also questions the homogeneity of the division level poststratum. These findings lend support to the accuracy of the adjusted count based on the synthetic method particularly within poststrata and block-to-block variance in characteristics and net overcounts and undercounts. Overall, 79 percent of the time the adjusted count is better than the census count. Nonetheless, this research 'flags' the need to be aware of State effects." (Estrada, page 19.) Given the fact that reapportionment depends critically on state counts, Estrada's conclusions raise a large flag in terms of accuracy.

[83] See Estrada, page 19; Wolter pages 7–8; and Ericksen *et al.*, pages 20–21.

[84] Estrada, page 18.

[85] Recently, Mississippi's proposed redistricting plan was overturned by the Department of Justice for failure to use block-level data.

much higher levels." [86] As an example he notes that "the [matching] effectiveness rates varied from a low of 87.2% in Albany to a high of 93.49% in Kansas City. . . . [T]here was a significantly different level of success in Kansas City than in Albany. But why? The answer is that we do not really know." [87]

Special Advisory Panel Member Tarrance links the usability of adjusted counts for redistricting with the disruption the use of such counts would cause.[88] These arguments will be considered under Guideline Seven.

Wachter has serious concerns about the usability of these adjusted counts. I consider his concerns about state population totals and reapportionment under Guideline Three.[89] He does, however, present evidence that casts serious doubt on homogeneity within post-strata. Because "very little is known about local heterogeneity in census coverage," [90] he conducted simulations on 10 selected PES block clusters to determine the effect of an adjustment on both the improvement in the numeric level of the population at the district office level and the improvements in the shares or proportions of the population in a given district office. In other words, he considered both numeric and distributive accuracy. In Wachter's simulations, the level of the population is improved about twice as often as it is worsened by an adjustment. However, the shares suffer much more from the simulated adjustment. On average 59% of the office proportions are better, but the range over all the simulations shows anywhere between 39% to 78% improvements. Furthermore, in 7% of the simulation trials a majority of the districts are made worse. Now in any simulation, a true population for a block must also be simulated. Wachter argues that truth is chosen in his simulations so as to overestimate improvements achievable by an adjustment.

Wachter's evidence on heterogeneity is the only evidence that looks at actual behavior in the 1990 census and PES below the state level, and the only evidence that looks at the effect of heterogeneity on the shares of the population rather than the population levels. He states that his results are preliminary and need more work—but at least they are results that bear directly on the homogeneity issue. I find compelling his conclusion that "local heterogeneity is a serious problem for adjusting the 1990 census at district levels. My evidence indicates that a substantial portion, possibly a majority, of relative counts for district-size units can be made worse off by adjustment." [91]

Wachter made other efforts to measure block-to-block heterogeneity and district-to-district heterogeneity. These other attempts are inconclusive and neither support nor deny the homogeneity assumption, so, therefore, I did not consider them to weigh either for or against an adjustment.[92]

Heterogeneity and local variability pose a vexing problem for synthetic adjustment as GAO noted in their testimony.[93] In his article, Freedman makes this clear:

> Variability is a major obstacle to adjustment. Indeed, undercount rates differ from one geographical area to another, and from one demographic group to another. That is why synthetic estimates for small areas, based on demographic analysis, have not been widely accepted. However, adjustment by the DSE [Dual System Estimate] is unsatisfactory for the same reason. For example, one post-stratum consists of Hispanics—cross-classified by age, sex, and housing tenure—in central cities in the Pacific Division (California, Washington, Oregon, Alaska, and Hawaii). In round numbers, the 1990 population of the Pacific Division is about 40 million with 8 million Hispanics, 5 million of the latter being in southern California.
>
> Consider an adjustment for Stockton, a city of about 200,000 people in California's Central Valley, a 4-hour drive north of Los Angeles. The Hispanic population is about 50,000; there can be at most a few dozen Hispanics from Stockton in the PES [Post-enumeration survey], and a handful of gross omissions [persons counted in the "p" sample who were not in the "e" sample (census)] or erroneous enumerations [persons counted in the "e" sample (census) who were not found in the "p" sample]. No stable estimates could be developed from a sample that small. Instead, estimates for Stockton would be based on the adjustment factor for the whole post-stratum, the numbers being driven by PES data from southern California. The basic assumption: undercount rates for Hispanics are the same in Stockton as in Los Angeles. *There is no empirical evidence to support this assumption.* [Emphasis added.] And there is a similar problem for non-Hispanics. Indeed, adjustment factors for non-Hispanics in Stockton are driven by PES data on non-Hispanics in the whole Pacific Division. Apparently, Stockton's non-Hispanics are supposed to be like their counterparts in the north, while its Hispanics are taken to be southern. *Stockton is the rule not the exception.* [Emphasis added.] There are 39,000 state and local government areas to adjust; and only 5,000 sample blocks with PES data. *Most jurisdictions would be adjusted on the basis of data from elsewhere* [Emphasis added.]—and the synthetic assumption.[94]

None of the evidence I was given, other than Wachter's, confronted the measurement of this problem head on. The questions that remain unanswered are fundamental: What is the extent of residual heterogeneity within post-strata down to the county, city, precinct and block level, and what is the effect of that heterogeneity on the adjusted estimates both in levels and shares? Until this is known, the statement that the counts are usable for all census purposes is no more than an assertion.

Conclusions

I conclude that the considerations pointed to by Guideline Two tend to reject use of the adjusted figures and support use of the census counts. The adjusted figures—like the census counts—are consistent across all jurisdictional levels and of sufficient detail for all purposes. However, the adjusted figures do not appear to be of sufficient quality to be usable for reapportionment and redistricting. First, the distributive accuracy of the census counts is superior as concluded above in my review of the evidence on Guideline One. Furthermore, substantial evidence casts doubt on the homogeneity assumption underlying the entire synthetic adjustment methodology. Even if the tests discussed under Guideline One and based on the homogeneity assumption had proven favorable to adjustment, this evidence would weigh against adjustment. Instead, both considerations imply that the adjusted figures are not of sufficient quality to be usable for reapportionment and legislative redistricting. Thus, this Guideline weighs in favor of a decision not to adjust the census.

Guideline Three

The 1990 census may be adjusted if the estimates generated from the pre-specified procedures that will lead to an adjustment decision are shown to be more accurate than the census enumeration. In particular, these estimates must be shown to be robust to variations in reasonable alternatives to the production procedures, and to variations in the statistical models used to generate the adjusted figures.

---

[86] McGehee, page 32.

[87] McGehee, page 4.

[88] See appendix 3: Tarrance, V. Lance "Report to the Secretary of Commerce," Member, Special Advisory Panel, June 14, 1991, [hereafter Tarrance], pages 17–18.

[89] Wachter, pages 24–26.

[90] Wachter, page 26.

[91] Wachter, page 29.

[92] Wachter, pages 30–32.

[93] See the quotation from GAO in Guideline One.

[94] Freedman, pages 1233–1236.

Explanation

The Bureau of the Census will determine the technical and operational procedures necessary for an adjustment decision before the results of the post-enumeration survey are known. This procedure shall be chosen to yield the most accurate adjusted counts that pre-census knowledge and judgment can provide. The Bureau of the Census will then assess the components of systematic and random error in the procedure and it will assess the robustness of the estimates generated from that procedure. Various procedures and statistical models can be used to generate estimates of net overcounts or net undercounts and adjustment factors. This guideline specifies that a set of procedures for generating proposed adjusted counts will be determined in advance of receiving the 1990 post-enumeration-survey estimates. This guideline requires that these procedures be evaluated. These evaluations will identify other procedures and models that could be considered as reasonable alternatives to the chosen production process. These alternatives will be used to assess the accuracy and precision of the proposed adjusted counts. In addition they will be used to assess whether and by how much the adjusted counts could vary if alternative procedures were used.

Discussion

There are three questions raised by this guideline that have not already been dealt with in my conclusions about accuracy in the discussion of Guideline One:

(1) Were the procedures followed pre-specified?

(2) Were the estimates robust to production alternatives?

(3) Were the estimates robust to alternative statistical models?

Prespecification

The question of prespecification is difficult. No production of the complexity of the census or the PES can be completely prespecified. There are always unforeseen events that occur and that require modifications to the plan. In fact the procedures for the PES and for generating an adjusted count of the population were, broadly speaking, as prespecified. Even though there were several decisions, of some importance, made in the course of the estimation procedure, all were made solely by the career professional staff at the Census Bureau. The decisions reflected the best professional judgment of those career public officials vested with the responsibility for the census and the PES.

First, a decision was made not to combine DA with the PES to generate dual-system estimates. Second, there was a choice made of carrier variables to be used in the smoothing process. These variables help determine how the raw adjustment factors (published on April 18, 1991) are converted to the smoothed adjustment factors (published on June 13, 1991). Finally, in the smoothing process itself some observations which were either peculiar in their magnitude or their variance were treated specially. The Special Advisory Panel members were consulted in trying to deal with the difficulties encountered in the smoothing process.

Kruskal, Tarrance, and McGehee all raise concerns about the prespecification question. It is Kruskal's impression "that choice of the so-called smoothing procedures was profoundly based on PES results. One might indeed argue that such a choice has major merits, but it does not seem to me to follow the Guideline" [95] McGehee argues more strongly: "One's confidence is further eroded when—in an effort to explain unexpected results—the Bureau resorts to novel explanations, re-manipulation of the data, and a variety of other *ad hoc* techniques." [96] Tarrance expresses similar concerns: "Some procedures have been pre-specified but, as in all statistical operations, others have been suggested and/or adopted as the operations have been carried out. I have been concerned to note that a number of changes have been made in the last 18 months." [97] He also notes that "any attitude of 'if the numbers don't come out the way we think they should we can change plans' is diametrically opposed to what good government policy should allow. Furthermore, it is clear that the adjustment process is a statistical operation which has never been done before and there are many last-minute decisions being made." [98]

Ericksen, Estrada, and Tukey either find no problem with the prespecified procedures or do not mention it. Wolter notes that there were procedures in the enumeration that were changed late in the enumeration process that affected the PES; however, PES managers were able to cope with the changes in procedures. He also notes the decision not to combine the PES and DA and the smoothing decisions made during the PES process. He finds that each was treated with a high degree of professionalism.[99]

In any estimation process unforeseen difficulties will arise and no estimating system can be put on automatic pilot. The unsettling problem is that, as we will see below, the choices that occurred did make a difference in the outcome of the adjustment—differences large enough to change the implied apportionment of the Congress—and that different choices producing different results may have been made by other responsible individuals in the exercise of their best judgment. The enumeration process itself cannot be influenced in such a way. Any individual decision either has a tiny impact or is so distant from the final result (both in temporal terms and in statistical terms) that the decision maker does not know the import of his decision. This is simply not true of the types of decisions made here in the course of calculating PES count estimates. State counts were easily available to the persons deciding which smoothing method would be used. Although I believe that the decisions were made for sound professional reasons in the 1990 census, using these adjustment mechanisms opens the possibility for manipulation of future post enumeration surveys in ways that are unavailable in traditional census procedures. This weighs heavily against an adjustment of the census.

Robustness of the Results

I will combine the discussions of the robustness to alternative statistical methods and production methods in this section because they are for the most part intertwined.

One area in which statistical models could have an impact on the result of the PES is in the imputation of match status. As individuals from the PES are matched back to the census some cannot be definitively declared matched or unmatched, often due to missing data. The missing data were imputed to the unresolved cases and a match status was then assigned using a series of statistical models. The levels of missing data were sufficiently low that variation in these models made essentially no difference in the outcome of the PES (Studies P1 and P2). Here I concur with the Undercount Steering Committee judgment that the outcome is robust to the alternatives considered, although, as noted above, Wachter warns that unexamined assumptions underlie the statistical imputation models and, in

[95] Kruskal, page 4.
[96] McGehee, page 4.
[97] Tarrance, page 20.
[98] Tarrance, page 21.
[99] Wolter, page 9–10.

fact, the results could be sensitive to these assumptions.[100]

Wachter notes that the sensitivity of the imputation results to these unexamined assumptions, however, could have an impact on the apportionment of the House of Representatives that would be implied by an adjustment. He considers five alternative adjustment calculations: the smoothed estimates, the raw estimates, two of his imputation alternatives, and a fifth estimate that uses state adjustment factors based only on PES data gathered within that state. He finds that each method implies a different apportionment of the House, and eleven states either gain or lose a seat in at least one of the five alternatives. This instability in the results of the adjustment for the Constitutional purpose of the census argues strongly against an adjustment.[101]

The second area in which different methods could have affected the outcome is in poststratification. All the members of the same post-stratum receive the same adjustment factor. If post-strata are chosen differently then outcomes may be different. The Census Bureau investigated whether changes in the post-stratification by census division would change the results significantly by using an alternative post-stratification by state. This showed that three states would have had significantly different counts. It is also important to note that any variation due to uncertainty in post-stratification is not incorporated in the total error model.

A third area of concern is that of smoothing procedures. Smoothing is a technique that is used to remove some of the effects of random variability in the estimates of the adjustment factors for the 1392 post-strata, while preserving the meaningful systematic differences between subgroups. Since these adjustment factors are the results of a statistical process, they are subject to random variation. If you had taken a second sample the answer would be different. But some variation across the different poststrata is a result of real differences in behavior not simply random statistical variation. The point of the smoothing exercise is to remove the *random* variation while attempting to retain the real differences.

Smoothing involves three major judgmental decisions—the treatment of outliers, the variance pre-smoothing, and the choice of so-called carrier variables. We consider first the treatment of outliers. This is an extremely complex problem that posed great unforeseen difficulty for the Census Bureau. Let me start with a simple observation. When the final PES numbers were announced on June 13, 1991, a modified set of PES numbers was included as one of the alternatives considered as a possible set of final PES numbers but not selected. This set of numbers stood apart from the census and was closer to the selected method than the census. Thus it was a candidate for selection. This alternative, had it been chosen, would have implied a different apportionment of the Congress than the selected method. If the selected method were chosen and if the Congress were reapportioned on the basis of those numbers, California and Arizona would gain one more seat each and Pennsylvania and Wisconsin each would lose one seat compared to the census. Use of the modified PES estimates instead of the selected method would have resulted in a shift of only one seat—from Wisconsin to California. It is important to note that the only difference between the two methods is that, in the selected PES. 28 outlying variances out of 1392 variances were omitted from variance smoothing. In the modified version these 28 points were not omitted. Thus changing the treatment of only 2% of the points could have changed the allocation of one seat in the House of Representatives. I have included in Appendix 10 a list of State, county, and city populations under three smoothing schemes: the selected method, the modified method, and the raw adjustment without smoothing. Some of the sensitivities to smoothing choice are evident from these charts themselves. Let me highlight a few:

- The undercount rate for Arizona is estimated to be 2.8% under the modified PES smoothing scheme and 3.3% under the selected PES smoothing scheme.
- The undercount rate for Maryland is estimated to be 2.5% under the modified PES smoothing scheme and 1.8% under the selected PES smoothing scheme.[102]
- The undercount rate for the District of Columbia is estimated to be 5.6% under the modified PES and 5.0% under the selected PES smoothing scheme
- The undercount rate for Akron, Ohio, is estimated to be 2.2% under the modified smoothing scheme and 3.0% under the selected PES smoothing scheme.
- The undercount rate for Pasadena, Texas, is estimated to be 3.7% under the modified smoothing scheme and 3.0% under the selected PES smoothing scheme.
- The undercount rate for Miami, Florida, is estimated to be 5.4% under the modified smoothing scheme and 4.6% under the selected PES smoothing scheme.

The Census Bureau analysis emphasizes that the set of various population estimates derived from different smoothing methods are broadly similar in the counts they produce and, as a group, distinct from the census enumeration. I believe that, in fact, it would be difficult to choose on any objective statistical grounds among the host of alternatives the Census Bureau considered which do in fact produce different results for the Constitutional purposes of the census. As noted in the discussion of Guideline One, accuracy must be considered in terms of the distribution of the population not numeric accuracy. The Census Bureau analysis does not consider the similarity in terms of the population distribution of the sets of estimates or whether the variance inherent in those estimates, warrants the discarding of the census in favor of one of the particular estimates.

Wachter's analysis of the smoothing procedures that the Census Bureau used in developing the adjustment estimates also raises some serious concerns. He believes that "smoothing has turned out to be the most problematic part of the adjustment calculations," and that "the evidence leads me to fear that the smoothing has had more of an effect on the final adjustment than can be easily justified." [103]

As noted above, smoothing is a technique that is used to remove some of the effects of random variability in the estimates of the adjustment factors for the 1392 post-strata, while attempting to preserve the meaningful

---

[100] Wachter, pages 21–22.

[101] In connection with the loss function studies discussed in Guideline One, the Census Bureau compared the apportionment implied by the census to that implied by the so-called target population. They differed by two seats. The Bureau then considered 1000 random draws from the production DSE statistical distribution and compared the apportionment that would result from each draw to the target population apportionment. For 391 of the draws the production DSE apportionment did not differ from the target apportionment. For 567 of the draws there was a difference of one seat. For 42 there was a difference of two seats. This only shows that the PES estimator of apportionment differed from the target apportionment by 0.65 seats on average. It says nothing about the quality of the census, since the target is simply another adjusted estimate of the population, as the discussion of Guideline One demonstrates.

[102] If a state is estimated to have a greater than average (*i.e.*, 2.1%) undercount it gains proportionally from an adjustment. States below 2.1% lose. Thus the choice of adjustment method, had adjustment been used, would have determined whether Maryland was a winner or loser.

[103] Wachter, page 33 and page 34.

systematic differences. This is done using a technique called linear regression that "holds constant" attributes of the population we expect to be associated with low or high undercount rates in an attempt to isolate the random variation. The choice of the attributes to be "held constant"—also called carrier variables—is a matter of concern and will be discussed below. These regressions yield estimates of adjustment factors that supposedly have been purged of their random variability. Wachter characterizes these estimates as being "flattened." [104]

To calculate the smoothed factors one takes an average of the raw adjustment factor (before flattening) and the flattened adjustment factor—but a weighted, not a simple, average is used. For a particular post-strata, if you have observed a lot of random variability, the smoothed factor is chosen to be closer to the flattened factor—that is, the weight on the flattened factor is high and the weight on the raw factor is low. On the other hand, if the raw adjustment factor is fairly stable and does not show much random variation, you put more weight on the raw factor and less on the flattened factor when you calculate the smoothed factor. The smaller the random variation in a poststratum, the more the smoothed factor relies on the observed data and the less it relies on the regression estimate.

But there is another level of complication. The measures of random variation, called variances, are themselves subject to random variation and, as happened in this PES, the variances can be large and unruly. The variances themselves vary a lot. When there are large measured variances, the smoothed factors are closer to the flattened estimates and on the whole, you tend to get lower adjustment estimates. The Census Bureau decided to soften this effect by pre-smoothing the variances before smoothing the adjustment factors. So there are two levels of smoothing—first variances, then factors.

Wachter shows that "the effect of deciding to use pre-smoothed rather than unsmoothed variances is to raise many of the adjustment factors by several percentage points and raise some by more than six percentage points. The changes introduced into the adjustment factors are of the same order of magnitude as the sizes of the adjustment factors themselves. These are huge changes for a decision of detail." [105] The fact that the statistical artifice of variance smoothing is making substantial differences in adjustment factors is disturbing. As Wachter observes:

> The raw adjustment factors are at only one remove from the data, the PES fieldwork that is the real information we have. Assumptions go into their computation and they are subject to many kinds of random and systematic errors. Notwithstanding these limitations, there is a fairly direct link between people missed or miscounted somewhere in a sample block and a big or small raw adjustment factor. Smoothing the factors themselves involves operating at two removes from the data, importing more assumptions, but incorporating information about variability that comes ultimately out of fieldwork. Pre-smoothing the variances that go into smoothing the adjustment factors is at three removes from the data. It incorporates little, if any, further empirical information. It depends entirely on another set of assumptions.[106]

The fact that pre-smoothing makes so much difference reflects the irregular and variable nature of the PES data. The implication is that the assumptions underlying the statistical models being used are important determinants of the outcome of the adjustment calculation.

Wachter discusses at length the reasons for variance pre-smoothing, but one argument he made was particularly striking. The variance pre-smoothing essentially results in large variances being made smaller and small variances being enlarged slightly. This seems to be the opposite of what is desired. A large variance means that the adjustment factor is not well estimated—it is noisy—so when smoothing the factor you should put *more* weight on the so-called flattened factor. Decreasing the large variance means you put *less* weight on the so-called flattened factor. The opposite argument can be made for small variances. Therefore, variance pre-smoothing is arguably having a result exactly opposite from that intended by the smoothing process. In addition, because low adjustment factors tend to have small variances, pre-smoothing makes those variances higher and thus systematically discounts the evidence of low net undercounts.[107] In other words, presmoothing tends to artificially inflate already high undercount rates and artificially dampen already low undercount rates.[108]

Wachter cites three other problems with variance pre-smoothing: First, the variance smoothing is not directed at making covariances more accurate. In his view the motivation for presmoothing was heuristic. Second, there are no strong reasons for choosing among the many models available to actually smooth the variances. Third, the choice about variance pre-smoothing affects not only the adjustment factors but the total net adjustments for broad aggregates of the population. For example, the variance presmoothing changes the estimated net undercount in the West South Central region from 2.95% to 2.76%. In the East South Central, it changes from 2.43% to 2.68%.[109] Again, these are changes of a very significant magnitude given the size of the national net undercount.

The choice of carrier variables in the statistical regression procedures used to smooth the adjustment factors could have a significant impact on the outcome. The Special Advisory Panel commissioned a study by David Hoaglin to study this impact. This study is used extensively in the arguments of Ericksen, Estrada, Tukey, and Wolter. The conclusion was much the same as with the various treatment of outliers. The carrier variable choice made a difference, although in absolute numeric terms not a huge difference. The 13 models Hoaglin produces look roughly similar to each other and to the production PES estimates all of which are distinct from the census. The same is true if relative shares for the thirteen evaluation post-strata are computed—the various carrier variables produce results closer to the production PES estimates than to the census. No results are available at finer geographic levels (such as states, counties, or cities.)

---

[104] Wachter, page 35.

[105] Wachter, page 36.

[106] Wachter, page 37.

[107] Wachter, page 39.

[108] In their letter submitted on July 11, 1991, Ericksen and Tukey dispute Wachter's concerns over variance pre-smoothing and contend that variance pre-smoothing helped the accuracy of the adjustment. In his rebuttal letter, submitted on July 12, 1991, Wachter stands by his statements. Both letters are contained in Appendix 16. It is difficult to referee this dispute in the eleventh hour, especially since the lateness of the Ericksen/Tukey letter gave little chance for Wachter to prepare a detailed response. In checking with the Census Bureau, I have found that, in fact, the pre-smoothing operation was agreed upon in advance, but in mid-May difficulties were encountered in that operation. The Census Bureau consulted with the panel and Tukey offered several remedies that Hoaglin and Glickman refer to as "prescriptions in the spirit of a mustard plaster . . . not a tightly specified procedure derived from established statistical theory." (Appendix E of Ericksen, *et al.*, page 15.) Although Hoaglin and Glickman, seem to indicate that the choice among the three remedies should not have much effect on the ultimate smoothed estimates, one of the three shows exactly the phenomenon that concerns Wachter, of raising high variances and lowering small ones. (See Appendix E of Ericksen *et al.*, page 22.) In fact, the choice of the "mustard plaster" did have an effect on apportionment (see main text above). Finally as Wachter points out, there is disagreement as to what constitutes a reasonable alternative.

[109] Wachter, page 39 and Table 3.1.

Wachter's assessment of the carrier variable selection is that "the effects of variable selection are not negligible but they are not a central issue." [110]

Ericksen points out that the total error model shows that the effects of the PES biases on population shares for the 13 large evaluation post-strata are small. In addition, he contends that his examination of the two estimates in the June 13, 1991, press release, shows the state population shares to be stable for the states that would gain or lose seats if the House of Representatives were reapportioned on the basis of adjusted counts. His reasoning is that the adjustments are larger than one or two times the standard error.[111] The difficulty with his reasoning is that it only considers sampling variability and ignores whether the shares are robust with respect to alternative statistical and production methods.

Conclusions

I have previously concluded that the adjusted figures have not been shown to be more accurate than the census enumeration. That is all that is required under Guideline Three to conclude that the census may not be adjusted. There are, however, additional considerations under Guideline Three under which I also conclude the 1990 census should not be adjusted.

It has proved virtually an impossible task to prespecify the adjustment procedure. It is equally impossible to prespecify the Census procedure. However, in the adjustment procedure an individual or responsible group must make choices which have politically significant effects on the counts that can be transparent to those making the choices. This puts the counts at greater risk of being manipulated than the census. There is no evidence of unprofessional or political manipulation in the 1990 PES program.

The results of the adjustment procedure are broadly robust at an aggregate, national level. However, although various alternatives seem to distribute counts in roughly similar ways, small changes in methodology can move seats in the House. It is also true that small changes in the census enumeration can move seats in the House as well, but no individual involved in the enumeration process can predict how. That is not true for the decisions for adjustment that cannot be or were not prespecified.

One of the most problematic parts of the adjusted process was the bundle of statistical techniques contained in the smoothing process. These techniques relied heavily on statistical assumptions, resulted in large changes in adjustment factors, and may very well have led to an overstatement of the undercount. Thus, this guideline weighs in favor of a decision not to adjust.

Guideline Four

The decision whether or not to adjust the 1990 census should take into account the effects such a decision might have on future census efforts.

Explanation

The Decennial Census is an integral part of our democratic process. Participation in the census must be encouraged. Respect for the objectivity, accuracy, and confidentiality of the census process must be maintained. Accordingly, if evidence suggests that adjustment would erode public confidence in the census or call into question the necessity of the population participating in future censuses, then that would weigh against adjustment.

On the other hand, if evidence suggests that the failure to adjust would erode public confidence in the census and thus result in widespread disinclination to participate in future censuses, that would argue for adjustment. The extent to which a non-adjustment would be perceived as a politically motivated act, and thus would undermine the integrity of the census, should also be weighed in making any adjustment decision.

Discussion

There is no scientific or quantitative means by which we can determine with reasonable certainty the impact of a decision made in 1991 on human behavior and activities that will occur in the year 2000 and beyond. Indeed, this guideline merely requires that we consider the effects that our decision today might have on future census efforts. In my view, such consideration requires that we examine relevant information and draw upon past Census Bureau experience as well as common sense in making rational predictions about such effects.

The universe of "future census efforts" encompasses a wide variety of activities: the efforts of individuals in completing census forms and cooperating with enumerators; the efforts of state and local officials, civic leaders and special interest groups in supporting outreach programs, public awareness campaigns, and active involvement in counting their target populations; the efforts of Census Bureau workers in enumerating as many households as possible; the efforts of Census Bureau professionals in making judgments and decisions about procedures to achieve the most accurate counts possible and to ensure objectivity and integrity of the process; and the efforts of the Department of Commerce, which includes the Census Bureau, to ensure appropriate levels of funding from Congress to support its enumeration activities. Each of these activities affects participation in and coverage of the census. To the extent that we can draw on relevant data, observations, and experience, consideration of the effects of decisions to adjust or not adjust on each of these activities is appropriate.

Sources relevant to our considerations include a study by the National Opinion Research Council (NORC), [112] public comments on the adjustment decision,[113] comments on Guideline Four submitted by the members of the Special Advisory Panel,[114] and discussions with experienced Census Bureau officials. Based on these sources, it is my conclusion that there is greater risk of potential harm to future census efforts as a result of a decision to adjust than as a result of a decision not to adjust. A discussion of the possible effects on each of these activities follows.

Effects on Individual Participation

Recently, the Census Bureau commissioned a study by the National Opinion Research Corporation (NORC) to try to measure how an adjustment might affect future census behavior by means of a telephone survey of a representative national sample of 2,478 households.

Persons were asked to evaluate the likelihood that they would participate in the next census. Then they were asked how that likelihood would change if there were an adjustment and how that likelihood would change if there were not an adjustment. The results were paradoxical—both a decision to adjust and a decision not to adjust would decrease the likelihood of participation.

The survey shows that the adjustment issue is not high in public consciousness or well understood. Only one-quarter (23.4 percent) of persons said they had seen or heard anything about the census in the past few months. When probed about what they had seen or heard, only 14.1 percent spontaneously mentioned anything to do with adjustment, undercount or errors in the census count. When told that people are talking about whether or not to adjust the

[110] Wachter, page 41.

[111] Ericksen, page 3.

[112] See Appendix 11. National Opinion Research Corporation, The Potential Impact of Adjusting of Not Adjusting the 1990 Census, June 19, 1991.

[113] See summary of comments on Guideline Four in Appendix 8.

[114] Ericksen, page 3; Estrada, page 20; Kruskal page 4; McGehee, page 33; Tarrance page 23; Tukey, page 2; and Wachter, page 42.

results of the census to correct for errors in counting the population, 22.3 percent then recalled they had seen or heard something about this. Probing questions showed that only 4.9 percent understand the adjustment issue.

Prior to any discussion of adjustment, a total of 84 percent of those surveyed stated they were "extremely or very likely to participate" in the next census. After the discussion of adjustment, 75.5% were "extremely or very likely to participate" in the future if the census were adjusted, as compared to 71.3% if it were not. Thus, while these results indicate that intention to participate in future censuses is marginally higher if the census were adjusted than if it were not, there is less inclination to participate in the future regardless of the outcome of the decision. As NORC points out in its conclusions: "While large numbers remain very favorably disposed to participating in the next and future censuses, this intention is a very slippery, ephemeral and changeable one . . . subject to influence by factors like the adjustment decision or, more likely, from the controversy or fallout emanating from the events that follow that decision." The survey also indicates that, prior to any discussion of adjustment, 5.5 percent were "not very likely" to participate in the next census. A decision to adjust would result in 5.3 percent in the "not very likely" category. A decision not to adjust would result in 8.6 percent in this category.

It is unclear what this survey meaningfully demonstrates, other than confusion over what an adjustment is and the negative effect of the controversy over adjustment on the present perception of a person's likelihood of participation in future censuses. However, as the survey report emphasizes, the need to explain the issue of adjustment and its implications will necessarily outlive the survey and the adjustment decision itself, and the inability of the surveyors to explain the issues to those surveyed is certainly grounds for some caution.

The division of public opinion on the future effects of adjustment indicated by this survey is consistent with the division of opinion demonstrated by the public comments received by the Department. While some claimed that an adjustment would erode public confidence in the census and thus lower future participation, others claimed that a decision not to adjust would erode public confidence and thus lower future participation.

The explanation of this Guideline states that evidence of widespread disinclination to participate in future censuses as a result of a decision not to adjust would weigh in favor of an adjustment. Neither the public comments nor the NORC survey provide evidence that this will occur. Indeed, the NORC study indicates that a decision not to adjust would make only 8.6 percent "not very likely" to participate in the future, just 3.1 percent more than those who would be "not very likely" to participate in any event. Thus, while there would be some additional disinclination to participate, it would not be widespread.

The explanation of this Guideline also states that evidence that calls into question the necessity of the population participating in future censuses would weigh against an adjustment. A number of the public comments express concern that an adjustment would result in the perception that an individual's failure to participate would be compensated by an adjustment and thus lower participation. In light of this, I am skeptical rather than optimistic about the likely motivation of individuals to participate in the future if an adjustment were made. However, I do not find compelling evidence in either direction regarding the effects of a decision on future individual motivations.

Effects on Complete Count Efforts by State, Local, Civic, and Interest Group Leaders

A number of the public commentators, as well as Wachter [115] and Tarrance,[116] expressed serious concerns that an adjustment would negatively affect the efforts by state, community, civic, and interest group leaders who traditionally provide essential support in encouraging participation in the census. I share these concerns. Currently, it is in the interests of every governor, mayor, and interest group to help get their target populations counted. The Census Bureau works closely with such officials and groups for two to three years before census day. The efforts include mapping, address compilation, massive advertising campaigns, and public awareness activities. I agree with my advisors who believe that such cooperative efforts are absolutely critical to the Census Bureau's mission to conduct an actual enumeration of all persons residing in the United States on census day, and particularly critical in reaching the hardest to count. Like others, I am concerned that an adjustment will remove the incentive that these public officials and groups currently have to provide active support in achieving a complete count.

[115] Wachter, page 42.
[116] Tarrance, page 23.

Based on the public comments, it is clear that many public officials believe that an adjustment will correct specific errors they have identified in the count of their communities. With such mistaken impressions, it is unrealistic to expect these leaders to put census outreach efforts above the many other claims on their limited resources. As Wachter predicts, complete count committees, local advertising, celebrity appearances, and special programs to ensure more complete minority counts would be likely to suffer as a result of an adjustment.[117]

Senior officials at the Bureau, including the Director, agree with this assessment. At the same time, the Director believes that states and cities will still have an incentive to encourage participation in order to get the best possible city planning data. I find this unpersuasive in light of the numerous public comments received from local officials demonstrating a profound lack of understanding of the effects of an adjustment and a misplaced faith in its ability to correct particular problems they have identified in their communities.

I find no evidence indicating that local support would decrease as a result of a decision not to adjust the census.

Effects on Funding of Future Censuses

Tarrance [118] and Wachter [119] expressed concern that an adjustment would adversely affect the Department's ability to obtain sufficient funding for future censuses. I share this concern. The most expensive element of the census is the extraordinary effort to count the last five percent. With the illusory prospect of an adjustment to achieve a full count in congressional districts and states, it would simply be unrealistic to expect Congress to appropriate funds to the full extent necessary to complete an enumeration of the hard to count. Without the funds needed to complete an enumeration, the quality of census data, especially in smaller areas, would be jeopardized. There appears to be little risk that Congress would deny such funds as a result of a decision not to adjust.

Effects on Efforts by Census Enumerators

As Wachter recognized, the future effects of a decision to adjust could be most severe on those temporary workers who must actually conduct the enumeration process.[120] The difficulties

[117] Wachter, page 42.
[118] Tarrance, page 23.
[119] Wachter, pages 42–43.
[120] Wachter, page 43.

of hiring, training, and supervising the thousands of temporary census employees are well-known and well-documented. It is time-consuming, often tedious, and occasionally dangerous work that requires extraordinary diligence for less than commensurate pay. There is a real risk that, with an expectation of a correction through adjustment, the field staff would not have the same sense of commitment and public mission in future censuses and, as a result, careless and incomplete work would increase, thereby decreasing the quality of census data. These are the workers the Bureau depends on to collect the data from the groups that are hardest to enumerate. If these data suffer, the information lost at the margin is information that is especially important to policy development.

I am unaware of any concerns that census enumerators would be less motivated as a result of a decision not to adjust the census.

Effects on the Independence of Bureau Professionals and the Integrity of the Census

Senior Bureau officials as well as Tarrance [121], Wachter [122], and McGehee [123] have raised concerns about the potential for manipulation of an adjustment for partisan purposes. As Wachter recognized, adjustment may pose significant risk to the technical independence of Census Bureau professionals who have traditionally been free from external influence in the implementation of their mission.[124] A principal drawback of adjustment is the fact that a few technical decisions can swing the outcomes of apportionment, redistricting, and Federal funding allocation. Decisions that may be nearly equally defensible from a technical standpoint may have very different outcomes which can be known in advance of the decisions. Thus, adjustment opens the door to manipulation of the census for partisan gain. It would therefore greatly increase not only external scrutiny and second-guessing of Census Bureau professionals and prospective candidates for key technical positions, but also inevitably increase pressure to politicize these positions. This would impose an even greater burden on technical staff in their attempts to make scrupulously objective and fair decisions. These risks pose serious threats to the integrity and objectivity of future censuses.

[121] Tarrance, page 5.
[122] Wachter, page 44.
[123] McGehee, page 33
[124] Wachter, page 44.

Concerns have also been expressed in the public comments and by Wolter [125] that a decision not to adjust the census may be seen as politically motivated and therefore adversely affect the integrity of the census. While I recognize these concerns, I believe they are outweighed by the likely adverse effects on future census efforts from an adjustment.

Conclusion

Based on the information available, I conclude that an adjustment would adversely affect future census efforts to a greater extent than any adverse effects of a decision not to adjust. The evidence indicates that the controversy over adjustment is likely to have a negative effect on future censuses regardless of the outcome of the adjustment decision. I am concerned that an adjustment would reduce state and local support for future censuses, adversely affect the Department's ability to obtain appropriate funding for future censuses, adversely affect the quality of the work done in the future by temporary census enumerators who are essential in reaching the hard-to-count, subject the Census Bureau to partisan pressures, and create the possibility for political manipulation of future census counts. Thus, this guideline weighs in favor of a decision not to adjust.

Guideline Five

Any adjustment of the 1990 Census may not violate the United States Constitution or Federal statutes.

If an adjustment would violate Article I, Section 2, Clause 3 of the U.S. Constitution, as amended by Amendment 14, section 2, or 13 U.S.C. section 195, or any other constitutional provision, statute or later enacted legislation, it cannot be carried out.

Discussion

In addition to the technical and operational aspects of the census and the proposed adjustment which I have considered in connection with Guidelines One through Four, I have also considered the constitutional and statutory implications of an adjustment decision. In my view, neither the Constitution nor the relevant statutory provisions are themselves conclusive as to whether the proposed adjustment would be unconstitutional or unlawful because the *sine qua nons* of constitutionality and lawfulness and the propriety of adjustment are the same: the need for unambiguous accuracy of the adjustment methodology and data. Because analysis of the significant legal issues is thus dependent upon the statistical analysis, which itself mandated my decision on the substantive merits not to adjust, it was unnecessary to decide the legal issues. This Guideline therefore only served to verify, not determine, my decision.

[125] Wolter, page 11.

Constitutional Considerations

While not free from doubt, it appears that the Constitution might permit a statistical adjustment, but only if it would assure an accurate population count. See *Carey* v. *Klutznick*, 508 F. Supp. 404 (S.D.N.Y. 1980); *Young* v. *Klutznick*, 497 F. Supp. 1318 (E.D. MI 1980). By implication, then, a determination that the proposed adjustment would not discernably or reliably improve the accuracy of the headcount would raise uncompromisable constitutional concerns, inasmuch as adjustment would not be contributing to the most accurate count, but rather would be injecting additional uncertainty and error. Thus, while the Constitution might not, *per se*, bar an adjustment, the question of whether a particular adjustment is constitutionally valid can only be made after the final form of the adjustment is known.

This principle—that an adjustment must be consistent with the constitutional requirement of "enumeration," i.e., an accurate count free from politicization and equivocation—is also supported by the intent of the Framers that the census utilize verifiable methods which obtain an accurate population count. This goal of accuracy would not be met, to give the clearest example, by mere guesswork. The central question under the Constitution thus supports, though it did not determine, my conclusion; the need for verifiable methodology and unambiguous data are the modern-day requisites of what was explicitly desired by the Framers when they provided for an "actual Enumeration." That phrase commands for all time that what shaped the details of the very first congressional apportionment (there was then as yet no census)—guesswork and political deal-making—never would be permitted again.

As the discussion of Guideline One demonstrates, the evidence of improved accuracy resulting from the proffered adjustment methodology is at best mixed. That evidence is not sufficient as a matter of substantive merit and, derivatively, it also fails the test prescribed under the Constitution. While the essence of my decision not to adjust rests in the uncertainty of the proposed adjustment and the questionable nature

of the data produced, that very uncertainty and question mark the rough shoals of politicization that the framers mandated be avoided when they required "enumeration," that is, an objectively accurate count.

Census Act Provisions

The Census Act contains two provisions authorizing the Secretary of Commerce to use sampling to conduct the decennial census. See 13 U.S.C. section 141(a) and 13 U.S.C. section 195.

Section 141(a) provides in pertinent part:

> The Secretary shall, in the year 1980 and every 10 years thereafter, take a decennial census of population as of the first day of April of such year, which date shall be known as the "decennial census date", in such form and content as he may determine, *including the use of sampling procedures and special surveys.* (Emphasis added.)

Section 195 provides:

> *Except for the determination of population for purposes of apportionment of Representatives in Congress among the several States,* the Secretary shall, if he considers it feasible, authorize the use of the statistical method known as "sampling" in carrying out the provisions of this title (Emphasis added.)

While judicial opinion is unsettled on the question of whether adjustment violates section 195, the majority of courts considering this issue have ruled that section 195 permits an adjustment if the adjustment method makes the census more accurate. See *Cuomo* v. *Baldrige,* 674 F. Supp. 1089 (S.D.N.Y. 1988), *Carey* v. *Klutznick,* 508 F. Supp. 404, at 415 (S.D.N.Y. 1980); see also, *City of Philadelphia* v. *Klutznick* 503 F. Supp. 663 at 679 (E.D. PA 1980); *City of New York et al.* v. *United States Department of Commerce. et al.,* (S.D.N.Y. 1990). *But see Orr, et al.* v. *Baldrige, et al.,* U.S.D.C., S.D. Ind., No. IP 81–604–C, July 1, 1985. Even assuming that the statute does not *per se* prohibit an adjustment, not all forms of adjustment would be sanctioned and the legality of the adjustment could only be determined after the form of adjustment is chosen. Thus, as with the constitutional issues, the analysis of the statutory issues cannot be separated from the analysis of the accuracy of the chosen adjustment method. Because the evidence of improved accuracy from an adjustment is insufficient, the standard articulated by the majority of these courts is not met. While this legal conclusion was not dispositive, it affirms my decision not to adjust based on the substantive merits.

Conclusion

The question whether the chosen method of adjustment would violate the Constitution and federal statutes depends upon the substantive analysis of whether accuracy of the census is improved by an adjustment. Because there are other compelling substantive reasons not to adjust, legal considerations did not provide a basis for my decision.

Guideline Six

There will be a determination whether to adjust the 1990 census when sufficient data are available, and when analysis of the data is complete enough to make such a determination. If sufficient data and analysis of the data are not available in time to publish adjusted counts by July 15, 1991, a determination will be made not to adjust the 1990 census.

Explanation

It is inappropriate to decide to adjust without sufficient data and analysis. The Bureau will make every effort to ensure that such data are available and that their analysis is complete in time for the Secretary to decide to adjust and to publish adjusted data at the earliest practicable date and, in all events, not later than July 15, 1991, as agreed to in the stipulation. Note, however, that the Department and the Bureau have consistently stated that this is the earliest possible date by which there is a 50 percent chance that an analysis could be completed on which a decision to adjust could be based. If, however, sufficient data and analysis of the data are not available in time, a determination will be made not to adjust the 1990 Census. The coverage evaluation research program will continue until all technical operations and evaluation studies are completed. Any decisions whether to adjust other data series will be made after completion of those operations.

Discussion

In order to evaluate the quality of the census and the post-enumeration survey, the Census Bureau conducted an extensive and ambitious research program designed to provide timely information on which to base a decision by July 15, 1991. Due in part to some unexpected anomalies in the data, progress on the evaluation was delayed in the final critical weeks, leaving the Bureau little time to complete its analyses. These pressures may have affected the quality of the research, and there is still much that we do not know about the quality of the PES and the adjusted counts relative to the enumeration. Nonetheless, based on the record available, I believe there is sufficient evidence to make a decision on adjustment.

The Census Bureau has done a remarkable job of condensing into a few short months a challenging evaluation program that was comparable to a multi-year research program for the 1980 census and the 1987 test of adjustment-related operations. The Census Bureau produced highly technical research on a very tight production schedule, using tools that were on the cutting edge of statistical theory and survey methods. The dedication, professionalism and hard work of Census Bureau staff under often intense pressure is truly commendable.

Although sufficient data are available for me to decide the adjustment question, it is important to note that because of the court imposed deadline for the decision, the analyses of the data are far from compete. All parties involved in the decision making process have expressed a desire for more time to digest and analyze the voluminous material created by the research program. I am particularly concerned about problems in data quality and analysis that were revealed, or occurred, in the final weeks before the decision.

Good research requires a careful weighing of the evidence, especially when it is on the frontier of the science. When such novel research is to be used for such far-reaching policy purposes, it requires discussion with peers who have not been intimately involved with the details so that some perspective can be gained. It benefits from probing questions, from looking at the data from different perspectives, from the use of alternative models and from intense and independent professional scrutiny. The time schedule simply did not permit a full range of such activities.[126]

Before the release of the selected and modified PES numbers, an inconsistency was found in the calculation of the margins of error upon reviewing the proposed release in its penultimate form. This was not a subtle error, but one that should have been caught by a careful cross-checking of the tables. After being informed of the inconsistency, the Census Bureau began work to discover its source. Fortunately, no fundamental error had been made. However, the release was delayed by almost two weeks, setting back an already tight schedule in the last critical weeks of evaluation. Such errors were the result of too much work being compressed into too little time. To its credit the Census Bureau worked hard

[126] Kruskal makes a similar point (Kruskal, page 6) as does Tarrance (Tarrance, page 27).

to find the error, fix it, and ensure that accurate data were released.

Later, in reviewing the work of the Undercount Steering Committee, fundamental questions were raised about measurement of the relative accuracy of the census and the PES. The loss function analysis was found to be unconvincing. The Census Bureau was therefore asked to revisit parts of its work. As a result of these questions, the Bureau staff found an error in the calculation of the loss functions. Correction of this error changed the number of States for which the census counts were judged more accurate than the adjusted figures from 11 to 21—a substantial and significant increase.[127]

An Addendum to the Undercount Steering Committee report was filed on Thursday, June 27, 1991. In section 4 of that addendum, which is included as appendix 5, the Undercount Steering Committee states the following:

> Given this new information, the Undercount Steering Committee members reevaluated their positions regarding the report issued on June 21, 1991 * * *
>
> The new information added uncertainty to the decisions of the majority, but their overall conclusions were not changed. In addition, particular sections of the report present representations of committee opinions that are now weakened by the new information. The sections of the report most affected by these new data are:
>
> The statement on page 6 of the report that 39 of 50 States are made more accurate by adjustment would be changed under the new loss function analysis; and
>
> Page 4 of the report summarizes the conclusions of the committee regarding Guideline One. The summary indicates that the majority of the committee relied on the loss function analysis that showed that a large majority of areas were made more accurate by adjustment. This is a stronger statement than the position now held by many of the committee members.
>
> In conclusion the overall committee position has not changed regarding adjustment, but has been weakened somewhat. These new data also underscore the points raised in report's findings on guideline 6 (see p. 12–13). When additional information, as the data presented above, becomes available, the committee acknowledges that it may strengthen or weaken its conclusions. On June 21 the committee judged that further analysis would be unlikely to change its conclusion. The majority stands by its original conclusion while acknowledging that the ongoing work, had it been available by the date our recommendation was due, may have caused different "weighing" of the results.[128]

These eleventh-hour findings weakened a key piece of evidence favoring adjustment. Because of these two significant errors, my concerns about the sufficiency of data and the strength of analysis supporting adjustment were heightened.

A second example of the pressures of the schedule is that as of the afternoon of Thursday, July 11, 1991, just two working days before my decision would have to be announced, I received a communication from Ericksen and Tukey taking issue with some of the conclusions in Wachter's report. Although I understand that many of the issues surrounding adjustment will be debated for a long time to come, the fact that some of the members of the Special Advisory Panel feel it incumbent upon themselves to offer last minute advice reinforces my perception that a full professional airing of issues has not taken place. Wachter wrote a speedy response to Ericksen and Tukey which I received on Friday, July 12, 1991. But a last minute debate by letter is not the way to carry out the important dialogue required on these issues.[129]

Over the course of the next months and years the data will be studied, the models tested, the professional discussions joined. We do not know what will be discovered about the quality of the PES data and the models that led to the adjusted counts. I am sure that the Census Bureau will not compromise its richly-deserved reputation for thorough and careful research. We need those efforts to build toward a better census in the year 2000.

But the question is whether we should adjust the census based on the data and incomplete analysis that we have now. As Wachter notes, we must "strike a sensible balance between the need to reach closure and the need to check and study further." [130] The decision must be made on its merits.

Notwithstanding my concerns about the effect the July 15, 1991, deadline had on research efforts, I conclude that sufficient data exist to permit me to decide whether to adjust the census. I conclude that the data support a decision not to adjust. Among the facts that weigh against an adjustment are:

- The PES missed a significant number of persons whom we cannot locate. Thus we cannot judge whether the adjusted census is distributionally superior to the enumeration simply by putting back into the count those we can locate by the PES.
- At the most aggregate level, the PES would move the count of the population in the opposite direction for some demographic groups as compared to those implied by DA.
- There is no convincing evidence to suggest that the adjusted counts give a more accurate count of the distribution of the population across various levels of geography. In fact the evidence indicates the census counts probably yield a more accurate measure of the distribution of the population.
- There is no convincing evidence that homogeneity within the poststrata used in adjusting the census counts is a statistically valid assumption.
- There is evidence that the estimates of the population produced by adjusting the counts are sensitive to small changes in the estimation procedure and these have significant effects.

Thus I find that the evidence presented is sufficient to conclude that the counts should not be adjusted.

Conclusion

An adjustment to the census is a fundamental change in the way we count and locate the persons residing in the United States. I am deeply concerned that if an adjustment is made, it would be made on the basis of research conclusions that may very well be reversed in the next several months. That would be bad for the country and bad for the Census Bureau.

The results of the PES evaluation studies are not yet completely analyzed. Because of the compressed time schedule imposed by the July 15 deadline, the analysis has not been subject to the full professional scrutiny that such important research requires and deserves. To the Census Bureau's great credit, the statistical tools used to calculate and evaluate the adjusted counts are at the cutting edge of statistical research. But such cutting edge research is not tried and true—it requires more thorough scrutiny before it can be used to affect the allocation of political representation and Federal funding.

Nonetheless, the demands of good research must be weighed against the need for a timely decision. In time we may find a way of combining the PES and the census to create counts that better reflect the absolute levels and the distribution of the population. There are sufficient data and analysis to support a decision not to adjust.

Guideline Seven

The decision whether or not to adjust the 1990 Census shall take into account the potential disruption of the process of the orderly transfer of political

---

[127] As noted in Guideline One, these numbers are for the version of the analysis in which it is assumed that the measured variance is the whole story. As discussed there, the change is even more dramatic (from 11 to 29) if the true variance is assumed to be twice the measured variance.

[128] Addendum, page 6–7.

[129] Both letters are contained in Appendix 16.

[130] Wachter, page 46.

representation likely to be caused by either course of action.

Explanation

This guideline is intended to ensure that the factor of disruption of the process of the orderly transfer of political representation is explicitly taken into account as the decision is reached. For example, many states have pointed to adjustment as being disruptive to their redistricting plans. Likewise, members of some communities that are believed to have been historically undercounted contend that if the Census were not adjusted, this would disrupt the orderly and proper transfer of political representation to their communities. The inability to ensure accuracy of counts at local levels may result in politically disruptive challenges by localities to official census counts.

This guideline recognizes that the Decennial Census plays a pivotal role in the orderly redistribution of political representation in our democratic republic. The process used to generate the required counts must not be arbitrary either in fact or appearance. The Secretary is thus obliged to consider the impact of his decision on the fairness and reasonableness of that redistribution to all those affected. This guideline requires an explicit statement of how and to what degree adjustment or non-adjustment would be disruptive. Even though these are concepts that are not easily quantifiable, they warrant serious consideration in order for the Secretary to make a prudent decision on an issue that profoundly affects public policy.

Discussion

Among the primary purposes of the census are to provide the basis for the reapportionment of the House of Representatives and the drawing of new Congressional district lines within states. Census figures are also used by most states to redraw state legislative district boundaries, as well as by cities and counties in redrawing their own districts.

The Clerk of the U.S. House of Representatives has officially certified to each of the fifty states the number of seats allotted to the state for the 103rd Congress based on the census figures released on December 26, 1990. As of May 1991, some 20 States had already enacted either or both of their Congressional and State legislative redistricting plans. The U.S. Department of Justice is reviewing approximately one dozen of the state plans as well as those of many cities and counties to ensure compliance with the requirements of the Voting Rights Act.[131]

If adjusted census counts were issued, Congress would have to decide whether to change the apportionment for the 103rd Congress which is to be elected in November 1992. If there were a decision to change the apportionment using the formula in current law, the Clerk would have to issue new certificates to the states advising them of the number of seats to which they would be entitled based on adjusted counts. If this change were made, the States of California and Arizona would gain one seat each and the States of Wisconsin and Pennsylvania would each lose one seat relative to the apportionment previously certified by the Clerk of the House.

It is unclear whether Congress would change the apportionment even if adjusted counts were chosen. The requirements of the statutes governing apportionment were fully met in January with the certification of the number of seats to each state. Thus, as noted in a number of public comments [132], additional action may be required on the part of Congress to change that apportionment. Whether, how, and when that action would be taken is for the Congress to determine.

It is important to remember, however, that the modern apportionment process was designed to be automatic. Once the counts were transmitted by the President to the Congress, the apportionment took place without legislative action. This design was intended to put an end to the blistering fights over apportionment that occupied earlier Congresses and, in fact, prevented reapportionment after the 1920 census, depriving citizens of a fair allocation of political representation throughout the nation for the remainder of the decade.[133] The adjustment of the Census might well create similar bitter disputes and paralyzing legal challenges over the apportionment of the 103rd Congress. The political implications of this are matters of substantial concern.

If the adjusted census were the basis for a reapportionment of the House, for the first time, the apportionment would not be determined solely on the basis of the number of persons within each State's border. This is due to the effects of cross-state groupings of post-strata in the PES on the adjustment process. For example, if the counts were adjusted, the certified population count for Delaware would depend on the results of the PES in Maryland, the District of Columbia, West Virginia, Virginia, North Carolina, South Carolina, Georgia and Florida. This is because Delaware is in the South Atlantic census division, and PES estimates are developed division-wide.

At the State level there is also likely to be confusion, disruption and extended litigation if the census figures are adjusted. Members of the Special Advisory Panel reported on extensive testimony they received from members of the National Conference of State Legislators in Baltimore, Maryland on June 28, 1990.[134] The testimony focused on the effects of an adjustment on the states' ability to accomplish redistricting in compliance with state-imposed legal deadlines. Witnesses were concerned that the electoral process would be paralyzed by the endless litigation which two sets of census numbers would be certain to provoke. Witnesses cited major problems with adjustment: costs and delays in drawing new plans, costs of additional elections, the need for costly special legislative sessions, time constraints, and charges of partisan tampering with census data. Based on the testimony, it is clear that adjustment would create serious disruption for at least a dozen states that have early redistricting schedules or constitutional deadlines. Some states have simply delayed starting the process until after the adjustment decision. As Estrada recognized, adjustment also would require modification of recently designed districts to meet one-person, one-vote requirements.[135]

Protracted legal battles that preclude redistricting in time for the 1992 elections would deprive minority groups and others the opportunity to realize and benefit from the gains achieved through demographic shifts during the past decade. The same pattern would likely occur in redistricting efforts for city and county elections, which have already begun in a number of areas. Moreover, the adverse effects of an adjustment on the accuracy of small area counts (as demonstrated in the discussions of Guidelines One through Three) would likely result in politically disruptive challenges by localities to adjusted counts.

Several public commentators, as well as Tukey,[136] noted that such disruption

[131] See appendix 12. Turner, Marshall, "Planning the 1990 Census Redistricting Data Programs," U.S. Bureau of the Census, [hereafter Turner].

[132] See the summary of public comments on Guideline Seven in appendix 8.

[133] See the discussion of this matter in Chapter Six of Margo J. Anderson, The American Census: A Social History. New Haven and London: Yale University Press. 1988.

[134] Tarrance, page 28 and Wachter, page 47.

[135] Estrada, page 23.

[136] Tukey, page 2.

was foreseeable at the time of the Department's decision to consider an adjustment and that anticipated effects should not be considered in making the decision. The fact that disruption could be anticipated does not mean that it should be ignored. Indeed, consideration of disruption as a factor to be weighed in the decision was legally upheld. Moreover, as Tarrance stated, "we would not be responsible stewards of the public trust if we do not understand that we are considering more than just a scientific statistical improvement of an imperfect government program." [137] Because the census is the basis for allocating political representation in our country, the public policy implications of adjustment, including resulting political disruption, had to be considered in reaching this decision.

The potential for disruption as a result of an adjustment must be weighed against any disruption that would occur from a decision not to adjust. There will inevitably be litigation resulting from a decision not to adjust that may also delay or disrupt redistricting. Some public commentators claim that the unadjusted census is itself disruptive because it does not ensure certain groups of their rightful claims on political representation and Federal funding. These claims rely fundamentally on the conclusion that the adjusted counts better reflect the distribution of the population. As explained in the discussions of Guidelines One, Two and Three, the evidence supports the contrary conclusion.

Estrada asserted that the public good is better served by focusing on the potential benefits to millions of persons rather than on the limited number of Congresspersons and state legislators who would be affected by a decision to adjust.[138] As demonstrated previously, the evidence indicates that millions of Americans may be harmed rather than benefit from an adjustment. Moreover, we must remember that the Congresspersons and state legislators who would be affected by an adjustment are elected by and represent millions of Americans in the political process.

Comments by members of the public and by Estrada [139] noted that an adjustment would result in more equitable allocations of federal funding to states and cities, a consideration which in their view must be weighed against any disruptive consequences from adjustment. Again, this claim assumes that the adjustment provides a more accurate distribution of the population across states and localities, an assumption which is not warranted by the evidence.

Moreover, it has been demonstrated that an adjustment of the census would have very little effect on the distribution of Federal funds. The study in Appendix 15 [140] shows that less than one fifth of one percent of Federal funds would be reallocated as the result of an adjustment. Twenty-one or fewer states would receive additional funds from an adjustment. Fewer than half of all jurisdictions would be allocated additional funds as the result of an adjustment. As the study demonstrates, those jurisdictions that do benefit would receive on average only $56 in additional funds per "adjusted" person.

Thus, even if the claim that an adjustment would more accurately (and thus fairly) allocate federal funds were valid, the adjustment would not result in significant shifts of those funds.

Conclusion

Any decision will result in some level of disruption through legal challenges. On balance, the record indicates that a decision to adjust would likely be more disruptive than a decision not to adjust. A decision to adjust would clearly cause disruption in those States that have early redistricting deadlines. The assertion that persons are denied their rightful claims without an adjustment assumes that the distribution of the population is improved by an adjustment. Based on the evidence, this assumption is invalid. Thus, this guideline weighs in favor of a decision not to adjust.

Guideline Eight

The ability to articulate clearly the basis and implications of the decision whether or not to adjust shall be a factor in the decision. The general rationale for the decision will be clearly stated. The technical documentation lying behind the decision shall be in keeping with professional standards of the statistical community.

Explanation

It is the responsibility of the government to have its critical decisions understood by its citizens. We recognize, however, that the degree to which a decision can be understood cannot alone dictate an important policy decision.

The decennial census is a public ceremony in which all usual residents of the United States are required to participate. If the census count were statistically adjusted, the rationale for that action must be clearly stated and should be understandable to the general public. If the decision were made not to adjust, the elements of that decision must also be clearly stated in an understandable way. It will be the responsibility of the Department of Commerce and the Bureau of the Census to articulate the general rationale and implications of the decision in a way that is understandable to the general public.

This does not require the Bureau or the Department to explain in detail to the general public the complex statistical operations or inferences that could lead to a decision to adjust. But, as with any significant change in statistical policy, the government has the duty to explain to the public, in terms that most can understand, the reason for the change. If the decision is not to adjust, (that is not to change) the public will be informed as well.

The last part of the guideline ensures that the methods, assumptions, computer programs, and data used to prepare population estimates and adjustment factors will be fully documented.

The documentation will be sufficiently complete for an independent reviewer to reproduce the estimates. These standards apply to the post-enumeration survey estimates, the demographic analysis estimates, and the small area synthetic estimates.

Discussion

The general rationale for this decision is clearly stated in the first section of this report. The technical documentation underlying this decision is in keeping with the professional standards of the statistical community. Thus the Guideline has been satisfied.

However, the Guideline could have been met if the decision had been to adjust. The Census Bureau has done a laudable job of keeping the public informed of the progress of the post-enumeration survey and the progress towards the adjustment decision. There is no doubt that the process of adjustment is complex and the statistical details of the process are fully comprehended by only a few individuals. Although I am sympathetic with these arguments, this would not have been an impediment to an adjustment. The general rationale could

[137] Tarrance, pages 2–3.
[138] Estrada, page 24.
[139] Estrada, page 23.

[140] See appendix 15. Murray, Michael, "Census Adjustment and the Distribution of Federal Spending," U.S. Bureau of the Census, May, 1991, [hereafter Murray].

have been clearly articulated. As Estrada notes, the public perception of the census "head count" is far removed from the actual process, [141] yet the general rationale for the census is well understood.

Conclusion

The requirements for this Guideline have been met. This Guideline does not weigh in favor of a decision either way since the requirements of this Guideline could have been fully met if the decision had been to adjust.

SECTION 3—SUMMARIES AND EVALUATIONS OF THE RECOMMENDATIONS OF THE SPECIAL ADVISORY PANEL

In this section I summarize the individual recommendations of each of the members of the Special Advisory Panel appointed to advise me on this decision, and the joint recommendation offered by Drs. Ericksen, Estrada, Tukey, and Wolter. After each summary I evaluate each recommendation.

Recommendation of Eugene P. Ericksen

Summary of the Recommendation

Ericksen recommends an adjustment. His argument relies substantially on a report co-authored by himself, Estrada, Tukey, and Wolter. He argues as follows: An adjustment will reduce the substantial error in the census and will correct for the differential undercount. The Bureau produced a demonstrably inaccurate census enumeration which can be fixed by means of PES estimates. PES estimates have been demonstrated to be both accurate and statistically reliable by evaluation studies of the 1990 decennial census. The racial differential undercount has again been demonstrated in the census, and the PES can correct for this clear and important bias.[1]

On Guideline One, Ericksen reports from his jointly authored analysis and other analyses that it is clear the adjusted count has been shown to be more accurate than the original enumeration. In Ericksen's view there is little doubt that the original enumeration is inaccurate. He states that the Census Bureau reported 13 million erroneous enumerations, 19 million omissions, and a PES net undercount rate of 2.1%.[2]

Ericksen says the basic flaw of the original enumeration is that it uses a method "designed for the well educated, middle-class family with reliable mail service." He argues that the method does not work for "those who do not read well, who live doubled up in an apartment, who live in drug infested neighborhoods with high crime rates, and who only occasionally receive mail." The procedure had such well demonstrated flaws that the 4.7 million undercount, and the 4.4% demographically estimated differential, was not surprising.[3]

Ericksen states that the PES was successful. The interviewing quality was high, imputation was minimal, and the matching error was very small. The evaluation studies suggest that the total error in the PES was minor. Correlation bias suggested that the PES underestimated the undercount, if anything. "The only reasonable conclusion is that the adjusted count is more accurate than the unadjusted count." [4]

On Guideline Two, Ericksen states that the adjusted data are consistent, complete, and of sufficient quality to be used for all purposes and at all levels for which census data are used. He cites the jointly authored report.[5]

On Guideline Three, Ericksen finds that "under any reasonable basis of comparison, the PES-adjusted enumeration is more accurate than the unadjusted enumeration." [6] The PES estimates are robust with respect to evaluation strata, and the effect of the PES biases on population shares is negligible. The estimates for the states whose Congressional delegation size might be changed by an adjustment are stable.

On Guideline Four, Ericksen says it is difficult to comment because of the lack of evidence. He interprets the available evidence from a National Opinion Research Center (NORC) study to suggest that most Americans would like to have the most accurate census and will trust the experts to make it so.[7]

Ericksen has no expert opinion on Guideline Five but notes that Jefferson lamented the lack of accuracy in the first census.

On Guideline Six, Ericksen feels sufficient data are available to make the decision now. Sampling errors for local estimates are reasonably small, and the PES evaluation studies indicate that bias is small.

On Guideline Seven, Ericksen admits having little comment. As a scientist he feels it is better to use improved numbers when available than to rush ahead and make errors.

On Guideline Eight, Ericksen believes that the results can be explained, and the technical documentation is in keeping with professional standards.

Evaluation of Recommendation

I agree the census had an undercount. I also agree that the evaluation studies demonstrated that the PES was well done. I do not agree, however, that the PES has the ability to correct distributional error. The grounds for my disagreement have been documented in the discussion of Guideline One.

I agree that the adjusted count, if more accurate, has been shown to be more accurate in a numeric sense at the national level. I do not agree that the adjusted count is more accurate in the distributional sense at lower levels of disaggregation. In addition, the erroneous enumeration and omission figures cited are Census Bureau estimates, which vary according to definition.[8]

The census used a variety of methods, including mail-out/mail-back, list enumerate, and list leave to fit different lifestyles. Class membership, education level, and reliability of mail service may explain some, but not all, of the census coverage problems. Recall that the personal enumeration censuses of 1940, 1950, and 1960 had even higher estimated undercounts. Thus, I disagree with Ericksen's notion that the census was "designed for the well educated, middle-class family with reliable mail service."

I do not agree that successful PES operations imply that the statistical manipulation required to go from its data to 4,830,514 blocks in order to produce a better count is a routine, automatic operation. I disagree that PES data, which are informative about the census, *can* be used to change the census in ways that make it distributionally more accurate.

I do not agree that merely because the Census Bureau can produce data that completely duplicate enumeration tables, that those numbers are of sufficient quality to be substituted for the census enumeration.

[141] Estrada, page 24.

[1] Ericksen, p. 1.

[2] Ericksen, p. 2.

[3] Ericksen, p. 2.

[4] Ericksen, p. 3.

[5] Ericksen, p. 3.

[6] Ericksen, p. 3.

[7] Ericksen, p. 3.

[8] The numbers used by Ericksen are estimates derived from all P-sample misses (19,171,290) and all E-sample "Erroneous Inclusions/Unmatchable." (13,154,639) While defensible, this is but one extreme definition. For example, it does not take into account the role of Census imputations. The matter of estimating these two components is a matter of disagreement among professionals. See, for example, the discussion in a Memorandum from Howard Hogan to Pete(r) Bounpane entitled "Gross Census Errors," July 2, 1991, Bureau of the Census on these issues. See the discussion of this issue under Guideline One above.

I agree that the PES adjusted enumeration may be more accurate numerically. I do not agree that it is distributionally more accurate. While the estimates are robust for evaluation strata, there is considerable doubt cast on their homogeneity with respect to post-strata relative to states.[9]

I appreciate Ericksen's comments on Guidelines Four and Seven, although I do not agree with them. I agree with his comments on Guideline Eight. I agree with his comments on Guideline Six, except that sufficiency of data in Guideline Six has nothing to do with substantive outcome, as Ericksen's comments about the size of sampling error would seem to imply.

Recommendation of Leobardo F. Estrada

Summary of the Recommendation

Estrada recommends in favor of an adjustment. Estrada first spells out a general rationale for his decision which is followed by an exposition of his reasoning for each guideline. Estrada relies on the paper co-authored by Ericksen, Estrada, Tukey, and Wolter.

Estrada's general rationale begins with the observation that the 1990 census is sufficiently flawed to require adjustment. In particular, the undercount rate increased from 1980, the census omitted the largest number of persons ever, historical undercount differentials between blacks and non-blacks persisted, and the black non-black differential actually increased from 1980 to 1990.[10]

Estrada states that the observed pattern of undercount is consistent with prior censuses. The Census Bureau efforts to overcome the undercount in the enumeration failed for a variety of reasons relating both to the character of the population and to the nature of the census operation itself. "While the Census Bureau was able to improve its internal management systems, the national dynamics that comprise the U.S. became more complex." [11]

Estrada argues that the differential undercount was the real cause for concern. He asserts that it occurred due to a number of problems in census processes. Flaws in the census operation included inaccurate mailing lists, non-delivery of census forms, a lower than expected mail return rate, inadequate interviewer and enumerator staffing levels, delay in district office closings, enumerator errors, enumeration by last resort, missing data, the inclusion of 2 million non-data defined persons in the count, lack of non-English language forms, processing errors, lost forms, race and ethnicity misclassifications, geocoding errors, and duplicate records.[12] District offices in the largest cities with the most heterogeneous populations suffered more from these flaws than others resulting in more last resort, close-out and non-data defined enumerations among non-Hispanic blacks and Hispanics.

Estrada states that the cumulative effect of all these problems is that the 1990 census needs adjustment.

Estrada describes the post-enumeration survey (PES) as a high quality process. He ascribes the high quality of the PES to, among other things, on-site listing of livable structures rather than reliance on mailing lists in sample blocks, experienced interviewers, a non-response rate of less than 1% and a proxy response rate of only 2.4%, relatively early interviewing to overcome the forgetting problem, successful tracking of the 8% of the PES who were movers, the successful evaluation program, and the fact of matching or resolving non-match cases for 98.3% of the 173,000 housing units surveyed.[13]

Estrada says that PES estimates of undercount follow known and expected patterns; *i.e.*, blacks higher than non-blacks, young males among minorities most often undercounted, the West division higher than other divisions, Hispanics highest rates of all. This attests to the "reasonableness" of PES undercount estimates and shows consistency with demographic analysis.[14]

Estrada claims the quality of the dual system estimates is sufficiently high to justify their use, according to the Hoaglin and Glickman sensitivity analysis among others.

Estrada says that adjustment methodologies improve the proportional distribution at all levels of census geography. He relies on the Tukey work and the work of other consultants that show that improvement at higher levels of geography improves shares at lower levels.

These conclusions by Estrada end the general rationale section of his recommendation The remainder of Estrada's recommendation focuses on each guideline.

On Guideline One, Estrada begins by reviewing the results of the Census Bureau evaluations of the PES, the so-called P-studies. The missing data studies (P1, P2 and P3) show that the rates of noninterview are low and the imputation for the primary population items was also low. Alternative means of imputing missing data did not affect post-strata. A Special Advisory Panel (SAP) analysis shows that post-stratum shares are minimally affected by eight alternative ways of handling missing data, with one exception. Given the small number of imputations required for the PES, alternative methods would have small effects on the outcomes.

Estrada says that the matching error studies (P7 and P8) confirmed that the high quality of clerical matching and matching of movers was performed successfully.

Estrada writes that the correlation bias studies (P13, P14, and P17) show strong correlation bias in the PES. Although for some this casts doubt on the dual system estimates, for him there is another side to the coin—"the undercount would be underestimated, particularly for minority populations. Whether the underestimation of undercount caused by correlation bias balances the biases toward overestimation of the undercount caused by missing data needs to be investigated, but the chances are they offset each other." [15]

Estrada states that other studies of data quality from the PES (P4, P5, P5A, and P6) show that the PES was not seriously impaired by problems of the quality in the reported census day addresses or fabrication.

Estrada says that those studies related to erroneous enumerations (P9, P9A, P10) show that erroneous enumerations were concentrated in particular evaluation post-strata. The census had higher rates of erroneous enumerations in minority areas and rural areas. Some significant changes would have occurred had matching of cases reported as erroneous enumerations been done by expert matching. On the census side there was a low error rate in matching, but more detailed analysis indicates that erroneous enumerations due to matching were more likely in two evaluation post-strata—non-minority areas outside the central city in the Northeast and West.[16]

Estrada claims that the study on late-late census enumerations (P18) shows that the addition of these data had an insignificant effect on the undercount rate. Similarly, balancing error was not a problem.

[9] See P12, and the discussion of Guideline Three above.

[10] Estrada, page 2.

[11] Estrada, page 3.

[12] Estrada, pages 4–6.

[13] Estrada, pages 6–8.

[14] Estrada, pages 8–9.

[15] Estrada, page 14.

[16] Estrada, page 16.

Estrada believes that the total error model (P16) indicates that errors introduced in the PES were small and tended to equalize racial differentials in the undercount.

On Guideline Two, Estrada states that a strong argument can be made that the requirement for local area accuracy can be satisfied by showing that adjusted counts are an improvement on the average for the principal uses of census counts. He claims it is appropriate to judge the adjusted counts at higher levels of aggregation than the block.

Estrada acknowledges that the Census Bureau study on heterogeneity (P12) shows mixed results with respect to the homogeneity assumption with respect to poststrata. "The research 'flags' the need to be aware of State effects [overwhelming poststrata effects]." [17]

On Guideline Three, Estrada acknowledges that the Census Bureau study on coefficients of variation (P15) showed that estimates of variances and covariances for smoothed and unsmoothed adjustment factors were larger than expected. However, he cites the Hoaglin and Glickman study as demonstrating the robustness and stability of the dual system estimators under different statistical treatments.

On Guideline Four, Estrada argues that if the Secretary adjusts using the best tools available, the reputation of the Census Bureau will be enhanced. The census process must incorporate adjustment as its final step. Estrada interprets the National Opinion Research Center (NORC) poll as indicating that the decision to adjust is slightly more likely to improve participation in future Censuses.

On Guideline Five, Estrada states that the innovation of adjustment is in keeping with prior Census Bureau efforts to meet the intent and spirit of the Constitution. The courts have already held that adjustment can be Constitutional.

On Guideline Six, Estrada states that "all the proposed studies have been completed, the data tables made available and the Census Bureau has had sufficient time to fulfill the concerns set out by [this guideline] in time for the Secretary of Commerce to make his decision." [18]

On Guideline Seven, Estrada states "Without denying the fact that there are State officials who feel imposed upon and elected officials (and potential challengers) who suffer from uncertainty as to when the boundaries of their districts will be 'fixed,' the actual consequences [of the census being adjusted and these figures not being available until July 15, 1991] are that a couple of Congressional seats will shift from one State to another; that delays will occur in redistricting, and that edges of many recently designed districts will have to be slightly modified to meet the one-person, one-vote requirements." [19]

Estrada says these disruptive consequences must be weighed against the fact that a census adjusted for deficiencies will provide a more equitable allocation of persons to each district, and a more equitable allocation for all other census purposes. The public good is better served by focusing on the potential benefits of adjusting the census to millions of persons rather than on the limited number of Congressmen and Congresswomen and legislative officials who will be affected by the July 15, 1991, decision to adjust the census and the subsequent release of adjusted numbers.[20]

On Guideline Eight, Estrada states there is an implicit assumption that the public understands the standard census methodology. However, their perception of what the census is—is far from the real census. Thus, both the real census and the reason for adjusting the census must be understood by the public. The public must understand the context of the PES in the census process. An informed public will accept the need to adjust if provided with concepts to understand the logic of the method.[21]

In conclusion, Estrada notes that the census has suffered from a persistent differential undercount. The evidence overwhelmingly demonstrates that the census count can be improved by adjustment. The PES adjustment factors have an advantage over demographic analysis in providing more specificity about the undercount. Adjusted counts will be more equitable and assure equal representation. Therefore, the Secretary of Commerce should adjust the census.

Evaluation of Recommendation

I do not agree that it follows that even were the 1990 census sufficiently flawed to require an adjustment, an adjustment is possible. The facts cited comparing 1980 and 1990 are a necessary, but not sufficient, grounds for considering an adjustment. A methodology must be available that will achieve a successful distributional correction.

I agree that the differential undercount is regrettable, and a cause for serious concern. I do not agree with Estrada that the flaws cited in the census are tied directly to that undercount. I agree that no matter how the differential came about, one would want to fix it *if one could.*

I agree that the PES was successful. However, I do not agree that the PES estimates followed all expected patterns. For example, in the discussion of Guideline One, above, serious questions are raised about its success in finding black males, and its "over compensation" for older females. In fact, PES results are frequently inconsistent with demographic analysis.[22]

I believe that the Hoaglin and Glickman study can be interpreted to show not robustness, as Estrada says, but that it can be interpreted to show that thirteen different models produce thirteen different sets of adjusted counts. These counts may have been close to one another, but not necessarily be an improvement over the census. Furthermore, as I noted in the discussion of Guideline Two there are other sources of variation due to statistical modeling.

I do not agree that the conclusions reached with respect to the Panel correlation bias studies are as clear as Estrada asserts. As Special Advisory Panel member Wachter suggests, the undercount may be underestimated by correlation bias effects not because of differential misses, but by differential erroneous enumeration rates when holding misses constant.[23]

I believe that Estrada's discussion of erroneous enumerations reaches the opposite conclusion from what the studies find: Differential erroneous enumeration rates by evaluation post-strata are a cause of concern, because they leave open the real possibility of differences between processing offices in how well the PES was carried out.

I agree that the total error model is experimental, but I disagree that the expression "total" is appropriate. Not all errors are included in it, only those errors that could be estimated on the basis of the PES. While the study of total error is encouraging, it is not yet dispositive with respect to the utility of the model.

Estrada acknowledges that P12 shows mixed results with respect to heterogeneity of post-strata. Thus, his assertion that requirements for local area accuracy are satisfied by "average improvement," and that only higher than block levels of aggregation need be considered, seems to me to contradict his acknowledging that local area

[17] Estrada, page 19.
[18] Estrada, page 23.
[19] Estrada, page 23.
[20] Estrada, page 23–24.
[21] Estrada, pages 24–25.
[22] See the discussion of Guideline One above.
[23] Wachter, pages 12–13.

accuracy needs to be satisfied. In fact, heterogeneity at the block level would mean that Guideline Two has not been satisfied.

As noted earlier, I believe that the Hoaglin and Glickman study can be interpreted as demonstrating a clear lack of robustness: Since accuracy at the block level is the goal, a process that allows thirteen different models to produce thirteen different estimates that differ only a little from one another, is not adequate. Differing a little at the high level of aggregation of the Hoaglin and Glickman work may mean differing dramatically at the block level.

I do not agree with Estrada's comments on Guideline Seven. The adjustment, as envisioned, will, in fact, not provide a more equitable allocation of persons to districts as he assumes. In my opinion, the lack of distributional accuracy is precisely why the adjustment is flawed as a correction for the census counts.

I do not agree that adjusted counts will be more equitable as Estrada claims in his discussion of Guideline Eight. In fact, they will not be more equitable distributionally, which is the criterion for determining whether an adjustment would improve the accuracy of the counts.

Recommendation of William Kruskal

Summary of Recommendation

Kruskal recommends against an adjustment. He uses the word "modification" rather than adjustment since the latter term suggests to him that "we really know how to improve the Census enumeration." [24] The primary reason for recommending against adjustment is that "we do not know with any confidence how to make such improvements . . . and we will not know in a relevant time scale." [25] Although "the proposed modifications are clever and technically interesting, the method turns on highly specialized assumptions and we simply do not know how robust the output results are against realistic errors in those assumptions." [26] The proposed modifications are complex, impossible to explain clearly for a general audience and their use is "likely to increase already existing apprehensions about manipulation and big brotherism in Washington." [27] The modified estimates might well introduce more error than they clear up, without anyone being aware of such an imbalance.

On Guideline One, Kruskal contends that there is no conclusive evidence that the modification removes more error than it introduces, and does not expect any convincing arguments anytime soon. The major gap in assessing comparative accuracy is the uncertainty about the "capture-recapture" model.[28] The implicit assumption of uniform capture probability is the most troublesome. Knowledge about the degree of output error caused by the non-factuality of this assumption "is just what we do not have, indeed cannot have, for the post-enumeration process." [29]

Later Kruskal notes that Guideline One calls for the highest professional judgment from the Census Bureau. "The highest level of professional judgment requires vigorous argument and discussion not only within the Bureau but in groups made up both of Bureau and outside statisticians and others. That vigorous and public discussion we have not had in nearly adequate amount." [30]

On Guideline Two, Kruskal's only comment is that synthetic adjustment is based on a simplifying assumption that is known to be wrong, which in turn throws great weight on the calculations of stability, given reasonable error structures.[31]

On Guideline Three, Kruskal's impression is that "choice of the so-called smoothing procedures was profoundly based on post-enumeration survey (PES) results," [32] which is not in keeping with the guideline. He questions whether "that in *major* respects the choice of procedure was made before the PES results were in hand," but time did not permit a full investigation on his part.

On Guideline Four, Kruskal feels the extraordinarily complicated procedures will undercut public confidence in the census. On Guideline Five, Kruskal has no comment. On Guideline Six, Kruskal believes that "timely data and analysis are not really at hand." [33] On Guideline Seven, Kruskal does not see how "this cuts in the present context." [34] On Guideline Eight, public explanation will be difficult because of the complexity and the choice of one of many such methods available.

Kruskal notes that the Guidelines "tilt against modification," but "that is hardly novel." [35]

Without resting his views solely on the guidelines, and instead on his "partly formulated and internalized professional criteria, along with [his] internalized civic standards," [36] Kruskal still recommends against an adjustment. He expresses concern about the large numbers of estimated counts and the inherent problem of putting together the millions of estimated differences between the count and the adjustment. He closes by noting that modifications that increase counts can, in fact, harm, by moving the proportions of the population in a given area in the wrong direction.

Evaluation of Recommendation

I agree that the census modifications lack robustness. Thus, Kruskal does not interpret the Hoaglin and Glickman studies as do plaintiffs' panel members. He recognizes that the adjustment may introduce more error than they correct without anyone knowing it.

I agree with Kruskal's criticism of the "capture-recapture" model upon which the DSE is based. He notes, in particular, that its assumption of uniform capture probability is most troublesome.

I agree with Kruskal's belief that there has not been an adequate vigorous and public discussion of the merits of adjustment. However, I disagree with his statement that the lack of such a discussion means we are not able to determine whether Guideline One is adequately met.

I disagree with Kruskal that, in terms of Guideline Three, there was no prespecification. He asserts that smoothing procedures were based on PES results. His comments implies a standard that would, in Guideline Three terms, preclude ever meeting prespecification requirements.

I agree with his comment that increasing counts can move proportions of the population in a given area in the wrong direction. That comment means that he, too, is concerned with the problem of distributive accuracy, and that he shares a concern about whether the proposed procedures deal with it adequately.

Recommendation of Michael McGehee

Summary of Recommendation

McGehee strongly recommend[s] that no adjustment be made to the census. There is no compelling evidence that suggests that the

[24] Kruskal, page 1.
[25] Kruskal, page 1.
[26] Kruskal, page 1.
[27] Kruskal, page 1.
[28] Kruskal, page 2.
[29] Kruskal, page 3.
[30] Kruskal, page 5.
[31] Kruskal, page 3.
[32] Kruskal, page 4.
[33] Kruskal, page 5.
[34] Kruskal, page 5.
[35] Kruskal, page 5.
[36] Kruskal, page 6.

PES [post-enumeration survey] will provide estimates that are any closer to the true population totals for the eight million blocks in the United States. Indeed, there is significant evidence to suggest that adjustment will move the population of many blocks further away from their true populations.[37]

Persons have always been missed in the census for a variety of reasons. Statistical adjustment is the most recent proposal to address the situation. [38]

McGehee states that adjustment numbers are estimates just like census counts: there is no way to determine which is closer to the true population, other than assumption and judgment. The evaluations of PES data "rested on pre-conceived assumptions of how the data would appear." [39] The results often fell outside the limits predicted from these assumptions. Rather than accepting the conclusion that the process is flawed, the assumptions were modified. He has no confidence in this reasoning. He refers to the problem in computing margins of error (variances) for local estimates as an example of this problem. "It is a strong indictment of the entire process, however, when evaluation procedures are not clearly understood by those using them * * *. The entire process has tended to produce more, rather than less, uncertainty." [40]

McGehee gives, as an example of the uncertainty created, the large difference in production matching effectiveness rates between Albany and Kansas City (87.20% v. 93.49%). Why this discrepancy exists is unknown and "no documented evidence can be presented which clearly explains this problem." [41] Adjustment proponents will argue that in the aggregate these problems are small and thus "the differences at lower levels should be overlooked because they become insignificant at the aggregate level." [42] McGehee disagrees, pointing to Guideline Two requiring accuracy across all jurisdictional levels. Furthermore, variation at the aggregate level, McGehee contends, is discounted by proponents by modifying the assumptions upon which the conclusions have been based.

"Decisions made during the DSE process, and the assumptions on which they stand, dramatically alter the adjustment results. A politically 'better' count cannot be defended if it is shown that the assumptions on which it rests are changeable." [43] Because of the widespread use of census figures, they must be defensible. The Bureau has maintained public confidence in its numbers over the years by "its meticulous approach to detail and its dogged adherence to maintaining the quality of Bureau data as the true standard." [44] Adjustment will undermine the public's confidence in this track record. A decision to adjust should be treated as political, and be forced to undergo the same Congressional scrutiny as other such decisions.

McGehee continues his argument by discussing the capture-recapture methodology. He uses an analogy to compare the PES to counting bears in a game preserve. He notes that the heterogeneity in game wardens' background and abilities, in the types of bears and their physical characteristics and in the terrain will lead to differences in how well the bears are counted. In similar ways, the enumerators' characteristics, the characteristics of the population the enumerator is counting and the environment in which the enumerator is working will all have effects on the outcome of the PES. These problems are compounded by the fact that PES records must be matched back to the census and the ability of matchers may be heterogeneous. [45] To identify the weight given to each of these variables, regression models are used to determine their individual effect. How these regression models are specified in the PES process is constantly changing. How to combine these variables into a larger number and how to compare various strata are issues of judgment on which individuals may differ. [46]

McGehee says that comparisons of data to the "correct" or "true" population are often made. The "correct" population is derived from a series of assumptions and thus results are simply theories. After reviewing the data, it is clear that the proposed adjustment does not meet the criterion of being usable across all jurisdictional levels nor is it robust at local levels to reasonable alternatives. The idea of using the PES to adjust the census is so complicated and so subjective, that no reasonable person can agree that it should be contemplated or that the process will be explicable to the general public. [47]

McGehee next turns to the issue of comparing the accuracy of the PES to the Census. Matching PES and census records is the key to assessing the relative success of the PES and the census in counting people. His "analysis shows that the PES fails to demonstrate a 'better' record of counting people than the Census. Indeed in many instances it cannot demonstrate that it did *as well* as the Census." [48] In support of his assertion McGehee presents a cross tabulation of census match codes by race and ethnic origin. He also does so for the PES. Although "time does not permit extensive analysis of this data," [49] he does note that twice as many Hispanics in the census left the race question blank as in the PES. More Hispanics identified themselves in the category "other" in the PES than in the census. "On a superficial basis, the results raise very significant questions whether adjustment will, in fact, yield greater accuracy than the census." [50]

McGehee states that the rationale for using the PES to correct the differential undercount rests on the assumption that as the black population increases in each block cluster, the PES will do a better job than the census in counting people. [51] It is appropriate then to compare the "best" and "worst" census and PES numbers within each block cluster and see how these comparisons change as the concentration of blacks increase over clusters.

McGehee argues that since errors occur in both the census enumeration and the PES survey, judgments had to be made as to whether it was correct to include them. These judgments are critical in determining the success or failure of the PES or the census. In those cases where judgments were made, one can get a range of estimates of quality by assuming that all judgments should have gone in favor of omission and, alternatively, all judgments should have gone in favor of inclusion.[52] Best and worst confidence level scenarios for the census and the PES in each block cluster are carried out. These comparisons are displayed by ranking the results in order of the proportion of blacks in the cluster, since research indicates "that as the percentage of black population within a cluster increases, the effectiveness of census coverage decreases." [53]

[37] McGehee, page 6, emphasis in the original.
[38] McGehee, page 2.
[39] McGehee, page 3.
[40] McGehee, page 4.
[41] McGehee, page 4.
[42] McGehee, page 5.
[43] McGehee, page 5.
[44] McGehee, page 6.
[45] McGehee, pages 8–10.
[46] McGehee, page 11.
[47] McGehee, page 12.
[48] McGehee, page 14.
[49] McGehee, page 19.
[50] McGehee, page 19.
[51] McGehee, page 20.
[52] McGehee, page 21.
[53] McGehee, page 26.

McGehee uses six graphs to present these results. "When comparing the best census scenario with the worst PES scenario one sees that the census does a dramatically better job of correctly counting people than the PES. . . . What is surprising, however, is the potentially dramatic performance shown by the census in those clusters where the black population is between 50% and 75%. Even more surprising is the very close correlation between the census and the PES in clusters where the black population is greater than 80%. In fact, the Census has a higher confidence level than the PES in those clusters where the black population is between 80% and 85%. This flies in the face—and graphically demonstrates the fallacy—of the argument put forward by the proponents of adjustment." [54] The PES does not necessarily outperform the census. Even if one accepts the midpoint between the best and worst PES results, the census exceeds this level and the PES does not outperform the census in clusters containing a large number of blacks.[55]

McGehee then turns to the guidelines. In his discussion of Guideline One, he finds the entire concept of adjustment on "the outer limits of statistical research." [56] The assumptions underlying the evaluations of the PES are so arbitrary and fluid that little weight can be attached to their assessments of PES quality. Therefore, Guideline One cannot be met since one cannot prove that the PES is better than the census.

On Guideline Two, he notes that variances between processing offices and evaluation strata are outside expected levels and at the district office level there was such variation it could not be reconciled. Adjusted numbers are inconsistent at the State, city, and subcounty level and suffer from serious quality concerns.[57]

On Guideline Three, McGehee asserts that the adjusted counts have not been shown to be more accurate than the census enumeration. The determination of quality is dependent on many assumptions and judgments.

McGehee says that the manipulation of assumptions in evaluation studies undermines confidence in all ongoing statistical data collection and therefore Guideline Four cannot be met.[58]

McGehee claims there remain legality questions about adjustment that need to be answered with respect to Guideline Five. On Guideline Six, McGehee states that sufficient data are available to suggest that the PES was flawed and the analysis of the data is insufficient to justify a decision to adjust the census.[59]

On Guideline Seven, McGehee finds that the mere fact of a possible adjustment has caused consternation and difficulties in state legislatures. The lack of consensus on the desirability and statistical feasibility of adjustment will result in extensive legal battles.[60]

Finally on Guideline Eight, McGehee asserts that the entire process is so complicated and difficult to understand, even by professionals, that a general rationale cannot be clearly justified. To the degree that the process is explained successfully people will become aware of the kind of manipulations underlying it and the integrity of the statistical process will be forever compromised. Adjustment is to correct an inequity, which is not a statistical problem but a political and societal problem that should be dealt with by the Congress.[61]

Evaluation of Recommendation

I agree with McGehee that the results of the PES fell outside expectations. The error variance around local estimates are an example of this problem.

I agree with McGehee's citing large differences in production matching effectiveness between processing offices as indicators of uncertainty rampant in the PES data. However, evaluation studies of the PES have not found the kind of systematic effect alleged.

I disagree that the link between the Bureau's credibility and its aversion to schemes that tend to devalue the census itself is a reason for avoiding adjustment.

I agree with McGehee's criticisms of the capture-recapture methodology. He raise issues not brought out elsewhere that cast doubt on its validity for use on human problems. I agree with his notion that characteristics of interviewer, interviewee, and setting interact to affect the quality of information, and find McGehee to persuasively elaborate the idea. I believe that McGehee's ideas support criticisms of Kruskal and others that the method is flawed fundamentally.

I disagree that if an adjustment were made it would not be explainable to the public. Since the decision not to adjust is just as complicated, this statement does not seem to have merit as an argument against adjustment.

Although I concluded that an adjustment would degrade the quality of the population distribution as compared to the census, I do not agree with McGehee's explanation of why the PES did not do as well as the census. He presents an analysis showing that, in a sample of block clusters, as the percentage of blacks within a cluster increases, the census actually performs better than expected. McGehee claims that this analysis casts serious doubt on the argument that *ipso facto* a PES based adjustment will necessarily reduce the differential undercount of blacks. I find his argument at best anecdotal and not compelling.

I agree with McGehee's conclusions that, on the basis of his analyses, arguments for adjustment based on Guidelines One, Two, Three, and Six are not adequate: The census remains more accurate than the PES; adjusted numbers are inconsistent at different levels of geography, and the quality of the PES is too dependent on assumptions, not facts and analysis.

McGehee argues on Guideline Seven that disruption is already occurring. This argument lacks support. He cites no evidence that adjusting or not adjusting will differentially contribute to disruption. Thus, I find that his arguments that this Guideline argues against adjusting are not relevant.

I disagree with his belief that the technicalities cannot be explained. Rather, I note that the process has been open, the Bureau has gone to great lengths to document its activities, so that there was no lack of ability to explain adjustment.

Recommendation of V. Lance Tarrance, Jr.

Summary of Recommendation

Tarrance recommends against an adjustment. He has chosen to concentrate on the public policy implications of a decision, not only because it is his area of expertise but also because he is "convinced that the impact of changes to the enumeration totals on the operations of our government—at the federal, state and local levels—would be disastrous." [62] Tarrance's lengthy introductory remarks are followed by a discussion of the guidelines.

Tarrance states that the perception that if the Bureau discovers how many persons it missed it should be an easy task to correct census results is

[54] McGehee, page 28.
[55] McGehee, page 29.
[56] McGehee, page 31.
[57] McGehee, page 32.
[58] McGehee, page 33.
[59] McGehee, page 33.
[60] McGehee, page 34.
[61] McGehee, page 35.
[62] Tarrance, page 1.

incorrect. In fact, there is no consensus on how to fix the counts among statisticians or other experts. Two Gallup polls—March 1990 and April 1991—show no consensus on including estimates of missed persons in the count. Whites were evenly split; non-whites preferred a synthetic adjustment.[63]

Tarrance says that more important than the statistical quality of the numbers is the public policy aspects of an adjustment. These include "the paralyzing difficulties that changing the numbers will cause in accomplishing redistricting . . . for all levels of the electoral system; the damaging perceptions that will be given to the public about the two different sets of numbers from the census; the troubling uncertainties surrounding even statistically acceptable numbers . . ." [64] Such policy difficulties should not be dismissed as many proponents of adjustment have done.

Tarrance asserts that lost in the debate fostered by adjustment advocates are the following points of decisive importance: (1) The adjustment process is complex, not well understood, without precedent and evaluations of it are judgmental; (2) synthetic estimates below the State level will never be more accurate than census counts; (3) the deadline of July 15, 1991, has not allowed enough time for adequate evaluation of the adjustment process or its product; (4) two sets of numbers may create "chaos" for the 1992 elections; (5) the trust in census confidentiality and the belief in the need to cooperate with the census will be further eroded; (6) resources may be denied to future census activities because "adjustment will take care of all problems" will be the expedient prevailing attitude; and (7) accepting adjustment will invite " 'inside manipulation' of numbers for political purposes." [65]

Tarrance says that "The adjustment process being used can produce an array of different results depending on the choice of assumptions and/or statistical methods employed. . . ." [66] Thus, the issue is not technical, but judgmental, as the decision calls for an assessment of the consequences of a decision. Whatever the decision, litigation will ensue, but a decision against adjustment "may be the beginning of a more reasoned look at the problem." [67] The Constitution says Congress shall determine how the census is to be conducted; therefore Congress should settle this issue, if at all possible, rather than the courts.

Tarrance quotes a statement made by co-chair Ericksen in 1980: "The undercount adjustment procedure needs to be statistically sound and politically credible," and goes on to state that the controversy has increased, in fact, and Ericksen's 1980 position is even more compelling today. Given the confusion and possibly paralyzing effects of adjustment, the best solution is not to adjust the census today, but to consider the proposal to adjust intercensal estimates as is done in Australia, Finland, and Spain.

On Guideline One, Tarrance first notes that statistical sampling only produces accurate results when sample sizes are sufficiently large, and for small jurisdictions this is simply not the case. Some small area counts will be made less accurate by an adjustment and the question is how we deal with such areas. There are a host of questions about tradeoffs among communities in accuracy that remain unanswered.

Furthermore, he points out that accuracy is a point of fundamental definitional differences between law and statistics: law needs certainty, statistics accepts a range of uncertainty about numbers it still considers accurate. "Any court settlement directing adjustment will *necessarily* require the arbitrary choice of numbers which have been derived from methods that statisticians would ordinarily hedge about. . . . It is paradoxical that those same interests who are faulting the Bureau of the Census for not having counted all persons are at the same time putting inordinate trust in that same agency to transcend the limits of statistical 'estimating!' " [68]

Tarrance argues that:

> The important fact that is buried in the mass of rhetoric about the need to correct inequities resulting from undercounting is that the numbers will undoubtedly be less accurate for many areas below the state level. The reality is that the adjustment process will not find those persons who were missed by the original enumeration and include them where they were not counted before. . . . Some correctly counted blocks could have persons added to their count; some correctly counted blocks could have persons deleted from the census count, and incorrectly counted blocks might not have any changes made to their numbers.[69] In addition, the post-enumeration survey (PES) is not able to handle all forms of counting errors with equal adequacy. Thus, misallocation can still occur even with adjusted numbers. Ultimately, "the final numbers are chosen from a range of possibilities that are dependent upon the choice of assumptions; there is a great deal of 'inside' judgment involved, and although [he has] no reason to doubt the experts at the Bureau of the Census who have had to make the hard choices, it is still troublesome that there is an opportunity for different results to be obtained by the use of different methods or assumptions." [70]

On Guideline Two, Tarrance states that a lack of usability for redistricting is a major deterrent to proceeding with adjustment, because of the conflicts having two sets of numbers will generate. "The realities of redistricting at the state and local level, combined with the possibilities for endless litigation, are such that it would be naive to believe that synthetic numbers will be usable . . . for the purposes of redistricting and reapportionment." [71] With two sets of numbers, redistricting plans will likely end up in court and the likelihood of "chaos" for the 1992 elections seems ever more probable.

On Guideline Three, Tarrance is most troubled by "the acknowledged fact that different methods using different assumptions produce different results." [72] As an example he notes that small numerical differences lead to large consequences in reapportionment and redistricting. "It is all too obvious that the procedures being used will not produce robust numbers and that it would be possible to obtain an array of population counts which could have very different effects upon apportionment." [73]

The requirement for pre-specification in Guideline Three concerns Tarrance, as some procedures were prespecified and some were not. In particular the decision not to combine demographic analysis with the PES was made by staff, in stream. This is an example of an attitude of "if the numbers don't come out the way we think they should, we can change plans" which is "diametrically opposed to what good government policy should allow. Furthermore it is clear that the adjustment process is a statistical operation which has never been done before and there are many last-minute decisions being made." [74] Tarrance

[63] Tarrance, page 2 and Appendices.
[64] Tarrance, pages 2–3.
[65] Tarrance, pages 4–5.
[66] Tarrance, page 6.
[67] Tarrance, page 7.
[68] Tarrance, page 13.
[69] Tarrance, page 13, emphasis in the original.
[70] Tarrance, page 16, emphasis in the original.
[71] Tarrance, page 18.
[72] Tarrance, page 19.
[73] Tarrance, page 19.
[74] Tarrance, page 21.

expressed uneasiness that "special interest pressure to adjust was pushing an incompletely researched or insufficiently tested statistical operation to a very shaky end".[75]

On Guideline Four, Tarrance states that a decision to adjust would have a far-reaching impact on future census efforts. Future censuses might be adversely affected as the Congress might well cut census funds, using the logic that an adjustment will fix the count anyway. Mayors and other local officials would question the necessity for their efforts on behalf of the census. The adjustment controversy might very well erode the already tenuous confidence of the public in the Census Bureau. The controversy surrounding the count should lead to imaginative ways to take the census in 2000, such as rolling samples, the "bare bones" head count, *etc.;* and legislative proposals immediately after the adjustment decision.

On Guideline Five, Tarrance states that Congress should determine how the census is to be conducted as required by the Constitution. Congress could also direct program solutions to resource allocation inequities.

On Guideline Six, Tarrance is convinced that the entire process has been rushed in an attempt to meet an arbitrary deadline. There has not been enough time for the evaluations. Given the controversy and that a general consensus has not developed, the adjustment should not be done without "the most exhaustive study and analysis of the data," which there has not been enough time to do.[76]

On Guideline Seven, Tarrance notes that the Special Advisory Panel met with representatives of the National Conference of State Legislatures. Technicians who must do the redistricting believe that they will be "paralyzed" by the "endless litigation" two sets of numbers will provoke if the census is adjusted,[77] although the very existence of two sets of numbers may be problematic. An adjustment would be most threatening to the creation of redrawn electoral districts for the 1992 elections.

Adjustment, according to Tarrance, will set a precedent for adjusting future censuses. He notes that one person miscounted in the PES represents from 500 to 1,000 persons that would be added or subtracted to develop adjusted numbers. The opportunity for, or perception of, manipulation to achieve desired ends will remain, but once adjustment is routine and not subjected to the scrutiny that it is now, the rigor of public examination to assure that manipulation does not occur will wane, and the risk, therefore increase.[78]

On Guideline Eight, Tarrance states that few people, even expert statisticians, really understand the process being used. He offers several examples of procedures and results of adjustment that are not well understood and states that it is impossible to articulate the complicated statistical procedures to the average person.

Evaluation of Recommendation

I disagree with the implication of Tarrance's discussion of public policy considerations that results of polls should play a substantial role for or against adjustment. I also disagree that if there is consensus that a particular adjustment would improve the counts and consensus that the adjusted counts are better than the enumeration, then an adjustment could be done based solely on that consideration.

I agree with Tarrance's point that there is support for not adjusting because of disruptive consequences for redistricting efforts.

I agree with Tarrance that the seven points of importance he cites, *i.e.,* complexity, lack of accuracy of synthetic estimates, inadequate time for evaluation, two sets of numbers leading to "chaos" for 1992, erosion of trust in census confidentiality, adverse consequences for funding future censuses, and the danger of inside manipulation, are valid expressions of concerns affecting the application of Guidelines One, Three, Six, Seven, and Eight.

I agree with Tarrance's discussions of lack of robustness which occur throughout the discussion. The point is made by him that judgment plays a substantial role in the choice of adjustment procedures. This is a flaw in the adjustment process pointed out in the discussion of Guideline Three, above.

I agree that Guideline One's requirements for accuracy are not met. The problem of misallocating people—even if one counts them correctly at a "higher" geographic level, is raised and documented. I agree that the arbitrariness of outcomes depending upon choice of assumptions is a fundamental weakness of adjustment.

I disagree that two sets of numbers will cause sufficient chaos to make either set not "usable" in Guideline Two terms. This is not the definition of usability intended by the guidelines. In fact, the effects of the numbers, if accurate and usable to the block level, should not play a role in the adjustment decision with respect to Guideline Two. This argument does not raise a bar to adjustment.

I agree that prespecification may be a cause for concern. However, because the prespecifications, such as the decision not to combine demographic analysis and PES results, were professionally done by career Census Bureau staff, I find that they impose no bar to adjustment according to this guideline.

I agree with Tarrance's assertion that adjustment will have an adverse effect on future censuses.

I do not agree that there has not been enough time for the PES evaluations.

I agree with the evidence as cited, including a meeting by the SAP members with representatives of the National Conference of State Legislatures, supporting, anecdotally, a prediction of endless litigation to be engendered by two sets of numbers, if an adjustment is made. I agree that there will be an increasing risk of future manipulation of the counts through adjustments if the precedent is set. This point is acknowledged in the discussion of Guideline Seven, above.

I disagree that the adjustment cannot be explained adequately, should it occur. I believe there is sufficient documentation to do so. I disagree with Tarrance's interpretation of the role of Guideline Eight on this matter.

Recommendation of John W. Tukey

Summary of the Recommendation

Tukey recommends an adjustment. He relies on the same report submitted by and coauthored by Ericksen *et al.* He argues that each and every one of the technical Guidelines are supportive of adjustment and the key Guidelines One and Three are indicative of an adjustment.[79] Tukey addresses the guidelines in the order given here.

On Guideline Four, Tukey states that a decision to adjust will enhance the Bureau's reputation and facilitate future operations, while a decision not to adjust may hinder future census efforts.

Tukey states that the questions raised in Guideline Five have been before the courts several times, and all decisions rendered permit adjustment.

On Guideline Seven, Tukey states that the Guideline must refer to aspects of orderly transfer of political representation that could not be

[75] Tarrance, page 22.
[76] Tarrance, page 27.
[77] Tarrance, page 28.
[78] Tarrance, page 29.
[79] Tukey, page 1.

anticipated in March 1990. There are no such aspects.

On Guideline Eight, Tukey states that the Guideline can easily be met. The technical documentation lying behind the adjustment decision is in keeping with the professional standards of the statistical community.

On Guideline One, in Tukey's professional judgment, the adjustments based on the post-enumeration survey (PES) have been prepared based on the highest professional judgment, and are more accurate, both as to numbers and as to shares, than the raw original enumeration

On Guideline Two, Tukey notes that, since the Bureau is preparing consistent and complete counts down to the block level, there is "no bar to adjustment." [80]

On Guideline Three, Tukey says that the Bureau has stuck to prespecified procedures. Dr. Robert Fay and consultants Drs. David Hoaglin and Mark Glickman have done a series of studies testing different statistical models that agree with one another and have proved to be good.

On Guideline Six, Tukey states there should be no questions raised about nonadjustment because of inadequate data by 15 July 1991.

Tukey ends with a post-script that notes that the existence of sensitivity of adjustment to reasonable choices should be no bar to adjustment, as long as it is small. The single prespecified procedure produces small sampling errors in comparison with post-stratum to post-stratum differences in adjustment factors to make it clear that adjustment provides smaller combined error than non-adjustment.

Evaluation of Recommendation

I disagree with the assertion that a decision to adjust will enhance the Bureau's reputation or facilitate future census efforts. In fact, other SAP members assert the opposite.[81]

I agree that Guideline Five is not a bar to a decision to adjust.

Tukey's interpretation of Guideline Seven, while unique, would not change the role this Guideline plays in the adjustment decision.

I agree Guideline Eight can be met.

I disagree that the analysis of Guideline One indicates that the Guideline has been met with respect to shares. Since the adjustment must clearly be shown to be superior to the census, controversy over this very important role played by census numbers indicates that the Guideline has, in fact, not been met.[82]

I disagree with Tukey's argument that Guideline Three has been met. In particular, I disagree with his interpretation of the Hoaglin and Glickman study, which he says supports the homogeneity assumption. As noted above, it can be used to support a conclusion that variance is a serious problem with the synthetic estimation model.

I agree that Guideline Six can be met.

I disagree that small differences between alternate sets of adjusted figures are no bar to adjustment, given the requirements to adjust to the block level with distributive accuracy.

Recommendation of Kenneth W. Wachter

Summary of the Recommendation

Wachter recommends against an adjustment. He "conclude[s] that the requirements for accuracy, state and local usability, and robustness articulated in Guidelines One, Two, and Three are not met by the adjusted counts. The broader considerations in Guidelines Four through Eight also, on balance, do not favor a decision to adjust. [He] therefore recommend[s] against adjustment of the 1990 U.S. Census counts." [83]

On Guideline One, Wachter concludes that the adjusted counts are not satisfactory. Although:

> evidence indicates that the adjusted counts are more accurate at the national level, the relative sizes given by adjusted counts are probably less accurate for a number of [S]tates and surely less accurate for a substantial fraction, possibly a majority, of local areas for which [c]ensus counts are to be used." [84]

As a preface to detailed sections on Guideline One, Wachter makes several pages of general observations:

The adjustment of a census is difficult as it is a matter of changing the counts for 6.8 million blocks. A post-enumeration survey (PES)-like survey is usually used to generalize up from sample totals to population totals; for such a use the absolute size of the sample rather than the fraction surveyed would limit the accuracy that could be achieved. The PES is used by the census to generalize down, which is a much more demanding process.

Three things must happen for the PES to be successful. The PES operation must be good, the people missed in the Census have to be reached by the PES, and the reasons why people are missed must be knowable so that one can extrapolate from the people and places analyzed to all the rest, for the PES to improve the census enumeration. The first has happened, the second has not, and the third is in doubt.[85]

The quality of the PES is high. There are problems and limitations but no disasters. Thus the first criterion is met.

A substantial portion of persons missed, net, by the census were not within reach of the PES. Discrepancies between estimates of national undercount between the PES and demographic analysis by age and sex for blacks and non-blacks cannot be explained away by plausible allowances for uncertainty. Half the black males who are missed, net, in the census are being missed, net, in the PES. There is no direct information on the distribution of these people from place to place.

As to the third criterion, the answers are not yet clear-cut. There is insufficient homogeneity at different levels of disaggregation for post stratum for the adjusted numbers to be usable. Erroneous enumerations are numerous and prominent in the adjustment picture. Block level data and district office data do not support the assumption of homogeneity.

Different smoothing procedures should lead to similar answers with respect to adjusted versus enumeration counts, but they do not: they lead to markedly different answers.[86]

Combining census and PES data produces results that are better than either alone only if we know enough about the precision and accuracy of each part. This is an empirical, not an *a priori*, question.

His personal experience with census enumerators and PES enumerators suggests that, contrary to common wisdom, census enumerators may very well have done a better job than PES enumerators in a significant class and number of cases.

[He] do[es] not believe that any highly aggregated index or loss function is appropriate for summing up overall accurary. It is informative to understand how much the outcomes of calculations with different versions of such aggregated indices differ. But the choice among them is not a scientific choice. Each such index involves implicit value judgments about different sorts of error. For example, each index determines whether a few large errors are more serious than a great many smaller errors. *Whether we agree with a particular tradeoff is a matter of personal and political values. It should not be disguised as science.*[87]

---

[80] Tukey, page 3.

[81] McGehee, page 6; Tarrance, pages 4–5.

[82] See the discussion in Guideline One above.

[83] Wachter, page 2 of cover letter.

[84] Wachter, page 1, emphasis in the original.

[85] Wachter, pages 1–2.

[86] Wachter, page 3.

[87] Wachter, page 5, emphasis in the original.

The census is the source of small-area data, so accuracy at that level has a special claim although some sensible balance of concern and perspective for level of detail is required.

In the first section devoted to Guideline One, Wachter considers national discrepancies between the PES and demographic analysis. There is a national undercount, although Wachter takes issue with the uncertainty intervals about the point estimates of the undercount. There is also credible evidence of a differential undercount. Although the evidence from the demographic analysis and the PES agree as to the existence of broad differentials, "the evidence as to the pattern by age and sex for blacks and non-blacks does not agree." [88] According to the demographic analysis, a high undercount rate for black adult males, aged 20–64 exists. This does not occur in the PES which means that *"a large portion of the people probably missed by the Census were also missed by the extrapolation from the PES that produced the adjusted counts."* He calls these people "unreachable." [89]

Wachter estimates the numbers of unreachable people to be large, perhaps half-a-million. Since nothing is known about their location, the huge numbers of "unreachable" people mean relative population sizes based on adjusted counts cannot be shown to be more accurate than those based on census counts at any subnational level.[90]

Wachter then turns to patterns in the estimates of net undercounts for post-strata. The patterns of adjustment factors for the 1392 post-stratum groups show regular patterns at higher levels of aggregation, but unexpected complexity when examined stratum by stratum, suggesting heterogeneity where there should be homogeneity. Analysis "for aggregates mask a large amount of diversity within groups, and the story of census coverage, at a level of fine detail, is more complicated than one would hope." [91]

Wachter then turns to the proximate determinants of net undercount. He finds that "erroneous enumerations account for a large portion of the variations in net undercounts across areas and post-strata." [92] Erroneous enumerations play a powerful role in determining the net adjustments to the counts, and this role is masked by smoothing adjustment factors which is probably unjustifiable.[93]

Wachter suggests that variation in erroneous enumeration could be the result of coverage improvement programs. The evidence that can be gleaned from comparing the cities of Detroit and Chicago is mixed. The main conclusion that can be drawn is that "erroneous enumerations are extremely varied. . . . [However,] lumping Detroit and Chicago together in the same post-strata, as the PES does, ignores sizable differences in coverage patterns." [94]

Wachter says that strong correlation between erroneous enumerations and omissions is insufficiently understood, even though it contributes substantially to the size of net undercounts. Since erroneous enumerations exceed omissions in a good number of post-strata, there will be a goodly number of downward adjustment. Thus *"people who themselves filled out their Census forms correctly may be 'minused out' of the Census to compensate for others who were erroneously enumerated"* to calculate an adjustment. "There may be no statistical objection to such a process. But on a human level it is offensive." [95]

Wachter asserts that there remain uncertainties in the demographic analysis, although it has been much improved.

Wachter states that the total error model does not mean all relevant errors for assessing the accuracy of a PES are included. Rather it addresses errors at the level of the evaluation strata only and, furthermore, treats them separately with no joint error structure. There is no simple way to generalize from the evaluation strata to small areas.

The approach is novel, pioneering and controversial. Thus, the "confidence intervals" around error components are not what statisticians usually mean by confidence intervals. The total error model actually estimates only a portion of the possible sources of error in undercount estimates. Components missed are of unknown magnitude. Stratification is applied inconsistently and some of the uncertainty estimates are themselves subject to large uncertainty. The total error model is too optimistic with respect to uncertainties attributed to imputation.

For Wachter, the main lessons drawn from the total error model are that the confidence intervals for most of the non-minority strata are compatible with zero net undercount, but the intervals for all the minority evaluation strata are not. The higher estimated undercounts are subject to high estimated biases.[96]

Several critical aspects of the total error model results are then discussed by Wachter, beginning with correlation bias or "catchability error." The correlation bias assumptions used are not realistic when applied to the PES. People stay out of the Census and the PES not by chance, but because they want to. Dual system estimation depends on chance mechanisms. There are many ways to allocate the twice-missed people. Whether the choice made is good is entirely speculative. How the measurement of variance in the total error model reflects correlation bias is not clear. It is better not to attempt any formal allocation of unreached people to local jurisdictions because of these problems.

Wachter next turns to matching and imputation studies. These studies of matching error give estimates of false non-matches that are too low by the very nature of their design. A small test on step-children illustrates the point that because matchers simply apply rules, they may miss true matches.[97] The effects of imputation may also be larger than the evaluation studies indicate. Wachter uses a sensitivity analysis to indicate the bounds on the effects of imputation. It shows that a great deal rests on the correctness of the assumptions in the imputation, but since these assumptions have not been examined, the measures of variance are too low.

On Guideline Two, Wachter sees "substantial obstacles to using adjusted data for Congressional reapportionment" and concludes that adjustment procedures are not well suited for coping with local heterogeneity in census undercounts. Firm conclusions cannot yet be drawn as to the extent of local heterogeneity and its implications for the accuracy of adjusted local counts.

Wachter shows by example that depending on how imputation is done seats could shift between States in a variety of ways. In estimating adjusted state population counts, adjustment factors based only on within-State data, rather than factors including across state data affect the distribution of Congressional seats as well. Among the five methods tried by Wachter, each apportionment was different and eleven states either gained or lost a seat relative to the census in at least one of the methods.

[88] Wachter, page 9.
[89] Wachter, page 7, emphasis in the original.
[90] Wachter, page 9.
[91] Wachter, page 10.
[92] Wachter, page 11.
[93] Wachter, page 10.
[94] Wachter, page 13.
[95] Wachter, page 14, emphasis in the original.
[96] Wachter, page 17.
[97] Wachter, page 21.

Wachter points out that there is acknowledged lack of homogeneity within post-strata. The issue is whether it is so severe to make adjustment locally infeasible. Very little is known about local heterogeneity. Experiments at the block level give ambiguous results with respect to the balance between improvements and worsenings of counts when adjustments are carried down to the block level. However, Wachter concludes that local heterogeneity is a serious problem for adjusting the counts at district office levels and that perhaps a majority of units could be made worse by an adjustment.

Wachter's experiments and analysis convince him that studies of local-level adjustment have "scarcely begun to scratch the surface" of the issue of how local heterogeneity has an impact on adjustment.[98] His block level analysis leads to more puzzles than answers.

On Guideline Three, Wachter finds that reasonable alternatives to one aspect of the smoothing model lead to significantly different adjustment factors and thus the adjustment factors cannot be considered robust. He finds that smoothing has been the most problematic part of the PES and that the smoothing has had more of an effect on the final adjustment than can be easily justified. The effect of deciding to use pre-smoothed rather than unsmoothed variances in computing smoothed adjustment factors is to raise many adjustment factors by several percentage points, some by more than six percentage points. The changes introduced into the adjustment factors are of the same order of magnitude as the sizes of the adjustment factors themselves.[99] Decisions about pre-smoothing make a large difference and so alternate methods leading to different outcomes seem equally reasonable. In fact, pre-smoothing seems to run the risk of "loading the dice." [100]

Wachter argues that pre-smoothing of variances changes variances in ways that are counter to what one ought to do: reducing large variances increases the weight assigned to empirically unstable factors; increasing small variances reduces the weight assigned to stable factors. In addition, the variance smoothing process is not directed at making covariances more accurate. Furthermore, the choice among regression models is arbitrary in the sense that there is no reason to choose among them, yet the results each set produces differ from one another substantially. Finally, smoothing affects not only adjustment factors, but higher level aggregations of data.

Wachter observes that the effects of the selection of variables for the regression part of smoothing are not negligible but they are not a central issue.

On Guideline Four, Wachter feels that an adjustment would reduce the stake that individuals, civic leaders and Congressional representatives would have in coverage improvement efforts. Adjustment would increase the political leverage of technical decisions and extra efforts to guarantee the Census Bureau's independence and objectivity would be required.

Wachter offers no guidance on Guideline Five.

On Guideline Six, Wachter states that sufficient data are available for a reasoned decision on adjustment.

On Guideline Seven, Wachter says that disruption is likely as a result of an adjustment, but this should not be decisive for the adjustment decision.

On Guideline Eight, Wachter sees no difficulty in meeting professional standards of the scientific community. The details of the adjustment decision tell against its understandability by the general public. Some dismay when an understanding of what adjustment really is should be anticipated, if the decision is to adjust.[101] Adjustment will have victims.[102]

Evaluation of Recommendation

I agree with Wachter's point that the PES, even if it yields results more accurate at the National level, doesn't improve the distribution of population over the results of the census enumeration totals due, in part, to "unreachable" people; among other factors.

I agree with the argument that a good PES is not a sufficient reason to adjust the census. I agree that Wachter's two other conditions are not met, *viz*, people who were missed must be reached, and why they are missed must be knowable.

I agree that Wachter's elaboration of the problem of correlation bias provides insight into why the adjusted counts produced from the PES may be distorted by correlation bias, and not simply underestimate the undercount. There are simply people who are unreachable, and determining why they are unreachable is an insoluble problem.

I agree with the analysis of discrepancies between the PES and demographic analysis.

I agree that the total error model does not include all, or necessarily even most, sources of error. I agree with the criticism that the confidence errors around the components of the model are speculative, and not uncontroversial among statisticians. Pointing out that higher estimated undercounts are subject to higher estimated biases casts serious doubt on the quality of these PES estimates.

I agree when Wachter states that the total error model does not mean all relevant errors for assessing the accuracy of a PES are included. I agree with him as he goes on to say, "Rather it addresses errors at the level of the evaluation strata only and, furthermore, treats them separately with no joint error structure. There is no simple way to generalize from the evaluation strata to small areas. The approach is novel, pioneering and controversial. Thus, the 'confidence intervals' around error components are not what statisticians usually mean by confidence intervals. The total error model actually estimates only a portion of the possible sources of error in undercount estimates. Components missed are of unknown magnitude. Stratification is applied inconsistently and some of the uncertainty estimates are themselves subject to large uncertainty. The total error model is too optimistic with respect to uncertainties attributed to imputation."

I agree with the discussion of Guideline Two that more work is needed to determine the homogeneity problem at the local level.

I agree with Wachter's conclusions with respect to robustness that interpret findings concerning the output from different models as raising questions about robustness at lower levels of disaggregation. In addition, smoothing is correctly identified as a significant factor affecting outcomes for higher level aggregations of data.

Recommendation of Kirk M. Wolter

Summary of Recommendation

Wolter recommends an adjustment. His analysis relies on the joint paper co-authored by Ericksen, Estrada, Tukey, and Wolter. The corrected counts, as required by Guideline One for an adjustment, are more accurate in both level and distribution at the national, state, and local levels.

Wolter finds Guideline One to be the pre-eminent guideline. His conclusion that the corrected counts are more accurate is based first on the observation that the post-enumeration survey (PES) is superior to the census by

[98] Wachter, page 30.
[99] Wachter, page 36.
[100] Wachter, page 37.
[101] Wachter, page 49.
[102] Wachter, page 49.

virtue of the design of matching operations and interviewer training and second, because a survey can be more tightly controlled than a census. The evaluation studies demonstrate that missing data, quality of Census day addresses, fabrication, matching, erroneous enumeration measurement, and balancing sources of error were controlled in the PES to very low levels. Correlation bias, while not so well controlled, is an error such that the PES estimates are still closer to the truth. Random error does not affect the utility of PES estimates.[103]

Wolter's rationale for preferring the adjusted counts includes four major points: (1) PES estimated undercounts agree with expectations and with demographic analysis; (2) the total error analysis demonstrates that corrected counts are more accurate for states, counties, and other similar areas; (3) corrected counts for evaluation strata, which are relatively homogeneous, offer even more improvement than they did for states, especially in comparing five minority with eight non-minority strata and central city versus non central city strata; and (4) if the stratum-level undercount rates are accurate, then the corrected counts for local areas must be an improvement on uncorrected counts.[104] This latter result is based on the Wolter/Causey paper that is appended to the coauthored report as Appendix G. Wolter also cites the plaintiffs co-authored report.

On Guideline Two, Wolter states that the bureau is capable of producing adjusted counts down to the block level, so the first part of the Guideline is satisfied. As to accuracy at small area levels, Wolter notes that, synthetic estimates of the kind used on the 1990 census can improve accuracy at small area levels so long as measured undercounts at aggregate levels tend to have smaller error than the original enumeration at aggregate levels. In support of his position, he again cites the Wolter/Causey paper. The Bureau's P12 study also offers evidence that the adjusted counts are superior to the census counts at the local level.

On Guideline Three, Wolter argues that the PES adjustment procedures were sufficiently prespecified to satisfy the guideline. The three instances where the procedures were not prespecified were "treated with a high degree of objectivity and professionalism."[105] The Hoaglin and Glickman report demonstrates that corrected counts are robust to variations in reasonable alternatives in the smoothing component of the overall PES process. The Census Bureau P1 study demonstrates that the PES undercount estimates are insensitive to differences in the manner of handling missing data.

On Guideline Four, Wolter states that "it is virtually impossible to say anything about the public's cooperation with the 2000 census."[106] The National Opinion Research Center (NORC) study indicates that the average American doesn't understand adjustment, plans to participate in future censuses, and that the adjustment decision, one way or the other, would have little effect. Other countries have instituted adjustment into their censuses with no adverse effect on public participation. Using the most accurate counts is the best way to handle the perception that the adjustment decision is a politically motivated act because Wolter believes that no matter what the decision is—it will be perceived as politically motivated.[107]

On Guideline Five, Wolter acknowledges that he is not a lawyer, but his understanding is that there is no legal ruling that stands in the way of an adjustment.

On Guideline Six, Wolter finds that the necessary data upon which to base the adjustment decision are sufficient, complete and available, and provide a sufficient basis for the adjustment decision.

On Guideline Seven, Wolter finds that the States have been alerted to the possibility of adjusted counts, and can deal with it. The Census Bureau analyses of misapportionment suggests that the original enumeration would misapportion seats more than the adjusted counts. Thus, not adjusting could be viewed as generating more disruption. Wolter is "unaware of any aspect of the 1990 correction process that would cause a truly calamitous disruption of the political process."[108] No part of the correction process has been arbitrary because scientific principles have guided the effort.

On Guideline Eight, in Wolter's view, there is a clear rationale for certifying the correct counts and the Bureau's documentation of the process has been satisfactory. The Bureau and the Department should be able to articulate clearly the basis for the adjustment decision

Evaluation of Recommendation

I do not agree that the PES counts are superior to the census counts. The four points of Wolter's rationale for believing the PES superior are flawed. Contrary to Wolter, PES undercounts do not agree with expectations, or the demographic analysis.[109] For example, the PES misses half a million black males which demographic analysis says are in the population. The total error analysis deals with numeric, not distributive accuracy. Thus, whatever it concludes about accuracy is not to the point of the form of accuracy which must be demonstrated.[110] The homogeneity assumption is in doubt.[111] There is not agreement on the inevitability of increased accuracy at lower levels, notwithstanding a certain degree of accuracy at broader levels.[112]

I do not agree that the synthetic estimate evidence in support of Guideline Two is clearcut, as Wolter states. In particular, P12 casts serious doubt on the homogeneity assumption necessary to a successful synthetic adjustment.[113]

I do not agree with Wolter's interpretation of the evidence with respect to robustness. I believe that the Hoaglin and Glickman report demonstrated that thirteen different models give thirteen different answers. An outcome of that kind is not robustness in the practical sense demanded by this guideline.

I agree that Guidelines Four and Five are no bars to an adjustment decision. On Guideline Six, I note that some panelists feel there is concern that census studies were not sufficiently analyzed in the time frame agreed to in the stipulation and order.

I do not agree that the Census Bureau analyses of misapportionment of Congressional seats are adequate.[114] I do not agree that there is clear consensus that the states can deal with adjusted counts.[115] In my view, while this does not bar adjustment, it remains a consideration to be reckoned with.

[103] Wolter, page 4.
[104] Wolter, pages 4–6.
[105] Wolter, page 9.
[106] Wolter, page 11.
[107] Wolter, page 11.
[108] Wolter, page 15.
[109] See the discussion in Guideline One above.
[110] See the discussion in guideline 1 above.
[111] See Appendix 2.
[112] Wachter, pages 2–3.
[113] See the discussion of distributive accuracy in Guideline One above.
[114] See the discussion in Guideline One above.
[115] See appendix 12.

Recommendation submitted jointly by Eugene P. Ericksen, Leobardo F. Estrada, John W. Tukey and Kirk M. Wolter

Summary of the Report on the 1990 Decennial Census and the Post-Enumeration Survey

The authors begin by considering the enumeration. The census differentially undercounts Blacks, Hispanics, Asians, and Native Americans. The Black undercount has been documented since 1940; the Hispanic since 1980. Differential undercounting is a result of the way the census is taken because it works best for "middle-class suburban" households and worst where living conditions are different. Undercount is strongly negatively correlated with the mailback rate.[116]

The authors state that the original enumeration of the population in 1990 experienced a staggering array of problems. The mail response rate was low, coverage differed between minorities and non-minorities, enumerators gathered less accurate information in cities than in other areas, and nonresponse follow-up operations had a high proportion of last resort and non-data defined responses. The difficulties in enumerating urban areas can be seen from the data. In large city offices 20% of all nonresponse followup was last resort or closeout versus 12% in small city/suburban offices and 11% in rural areas.[117]

The authors claim that last resort and closeout information is incomplete and often inaccurate. More than one-third of all last resort information and 44% of all closeout cases were estimated to be erroneous.[118] Re-enumeration of households originally enumerated by last resort or closeout showed serious errors in certain problem offices. In a national survey of 1,000 one-person households there was between a 20% and 25% error rate depending on the measure used.[119]

The authors say that coverage improvement programs, while adding people to the count, were frequently in error. For example, more than 630,000 of the 2.1 million persons added through vacant/delete either should not have been added at all or should have been added at a different place. More than half (53%) of the persons added to the count through the parolee/probationer check were estimated to have been added in error. Overall, the coverage improvement programs failed to do what they were supposed to—accurately add a substantial number of persons to the census count and the differential undercount remained after the programs had been completed.[120]

In addition to adding error to the count, the authors argue that the coverage improvement programs failed to find the estimated 19.2 million persons actually missed by the census. The "Were you counted" campaign and the Housing Coverage Check and Local review added only 200,000 and 300,000 persons, respectively, to the count. The low number of accurate additions left intact and possibly increased the differential omission rates by race and type of area that had already existed.[121]

The authors next turn to demographic analysis. Demonstrating through demographic analysis that a black non-black differential undercount exists for every census since 1940, the authors conclude that a black non-black differential undercount exists by virtue of demographic analysis in the 1990 decennial census.[122]

Next, the authors turn to the post-enumeration survey (PES). The PES is the mechanism designed by the Census to determine the extent of, and correction for, census error. The post-enumeration survey has demonstrated the differential undercount of the minority population and solved the major error of the original enumeration, which was the inappropriate shifting of shares of population from areas with many minorities to areas with fewer.

The authors state that the PES was a high quality survey. Completed interviews were obtained 99% of the time for the total PES sample, and for major geographic and racial subgroups. Proxy interviews accounted for 2.4% of the total sample, with little variation in this rate across subgroups. Only 1.5% of the P-sample were unresolved in the matching operation, and only 0.9% of the E-sample. There was little subgroup variation.

The authors use three criteria to evaluate the success of the PES: consistency with expectations of the distribution of the undercount (*i.e.* rates of omission and erroneous enumeration should be higher where census taking was more difficult) and the results of demographic analysis; the P studies (looking at missing data and the outcomes of rematch studies especially); and the possible shifting of population if net undercount rates were altered as a result of the P studies.

The authors state that PES results were consistent with substantive expectations especially when compared with demographic analysis.[123]

The authors' examination of P studies focused on four problems: The effect of variation in assumptions on how to treat missing data; problems due to matching error; problems with census day address misreporting and matching error for movers; and correlation bias. Assumptions about how to treat missing data had little effect. Because the numbers of movers were small, mover matching error had little effect. Correlation bias was a major source of error. Its effect tends to be to reduce estimated undercount. Evidence from evaluation poststrata research shows that adjustment increased the minority share of the nation's population by 0.8%, from 21.4% to 22.2%. The total error model showed a shift of 0.76%.[124]

The next major area considered by the authors was the smoothing of the adjustment factors. They consulted with David Hoaglin to evaluate the impact of the decisions on carrier variable choice, how to smooth variances and covariances of raw adjustment factors before calculating the regression, and how to weight individual observations when calculating the regression.

Hoaglin identified how to smooth the variances before using them to weight observations in the regression calculations and how to smooth the covariances before using them for the same purpose as key decisions.

Hoaglin fitted thirteen different regressions. The first nine were based on three strategies for smoothing variances and three strategies for smoothing covariances ($3 \times 3 = 9$); a tenth alternative was suggested by a Panel member; finally for comparison purposes he considered equal weighting of observations; weighting according to raw variances and covariances; and weighting according to raw variances, replacing the covariances by zero.[125]

After considering various alternative "stopping rules" for the "best subsets regression," Hoaglin chose a "back-2" stopping rule which uses apparently the best subset among those involving two fewer carrier variables than are in the set that minimizes the ratio residual mean square/residual degrees of freedom.

Hoaglin used two strategies to test whether the decisions had serious impact on the estimates: The first strategy used the difference in fitted values from each pair among the 13 choices and differences between the 13

[116] Ericksen, *et al.*, pages 1–2.
[117] Ericksen, *et al.*, pages 4–5.
[118] Ericksen, *et al.*, page 6.
[119] Ericksen, *et al.*, page 6.
[120] Ericksen, *et al.*, pages 7–8.
[121] Ericksen, *et al.*, page 8.
[122] Ericksen, *et al.*, pages 10–11.
[123] Ericksen, et al., pages 13–14.
[124] Ericksen, et al., pages 12–16.
[125] Ericksen, *et al.*, page 18.

and the Bureau's regression fit; while the second strategy used the reallocation of population shares among the 13 evaluative post strata.

Hoaglin stated that alternative smoothing models produced estimated population share gains for minorities that closely "surround the Bureau fit," ranging from 0.48% to 0.77%[126]

Next the authors considered errors for large and small areas. In looking at the differences in errors for large and small areas, they concluded that the total combined error increases as the size of the group decreases (*e.g.*, the combined errors for 5 million blocks will be larger than the combined errors for 1,392 poststrata), and consequently the improvement in amount due to adjustment would be nearly the same for larger and smaller groups—the improvement in percentage terms decreases, but does not change sign, as the groups become smaller.

The authors stated that since the expected CV for a sampling stratum is 1.4%, they were more likely to expect improvements for those areas where undercounts are especially high or especially low. It is these extreme cases where most of the benefit of adjustment is to be expected. Improvements in quite large areas thus prophesies improvements in very small areas, as well as in intermediate areas.

The authors' major conclusions are that error in the uncorrected census was very high; this error disproportionately affected Blacks, Hispanics, Asians and Native Americans; and the PES derived data can be used to correct the census and substantially reduce the differential undercount and improve accuracy at both national and local levels.

Evaluation of the Report on the 1990 Decennial Census and the Post-Enumeration Survey

I do not find the discussion of the quality of the census relevant. Guideline One stipulates that the census is the standard. Thus, irrespective of the flaws in the census, Guideline One precludes adjustment unless the adjustment is shown to be better than the census by convincing evidence.

I do not agree with the statements in discussions of the PES claiming that PES results were consistent with expectations when compared to demographic analysis is made. There were sizable, and unexpected differences between the PES and demographic analysis which indicate that a PES based adjustment would be inadequate. [127]

I do not agree with the interpretation of the Hoaglin materials. The authors' interpretation misses the point. The issue is not whether the thirteen different outcomes fluctuated around a Bureau estimate of "truth" derived from the PES and are thereby defined as demonstrating sufficient robustness. The very fact of such a variety of outcomes is precisely the lack of robustness that is of concern when using a model based synthetic adjustment at a low level of geography.

The authors state that the expected CV for a sampling stratum was 1.4%. The expected CV was .7%.

I do not agree that PES derived data can be used to correct the census and substantially reduce the differential undercount and improve accuracy at both national and state levels. [128]

SECTION 4—DECENNIAL CENSUS PROCEDURES

In this section I provide documentation for the procedures used to conduct the decennial census, the post-enumeration survey, the evaluation of the post-enumeration survey, and the evaluation of the demographic analysis. Additional information on the post-enumeration survey evaluation program and demographic analysis will be found in appendix 3.

1990 Census of Population and Housing: The Bicentennial Census of the United States

Planning for the 1990 Census began in 1984, with planning activities, testing, and preparatory operations occupying the remainder of the decade. Data were collected in 1990, and, as required by law, State population and apportionment totals were delivered to the President on December 26, 1990. The total population count transmitted to the President was 249,632,692, composed of a resident population of 248,709,873 and an overseas population of 922,819.

The Census Bureau was also required by law to deliver redistricting counts and maps to State redistricting officials no later than April 1, 1991. This was done. While the Census Bureau met its two legal mandates for the delivery of apportionment and redistricting data—two of the most important uses of census data—the 1990 census is not considered completed until all planned census data products have been released. Final products will be released in 1993.

The 1990 census involved enumerating 249,632,692 people in more than 100 million housing units, and collecting a full range of characteristics about each person. Extensive planning and preparation, the successful recruitment and employment of hundreds of thousands of temporary census workers, and an automated management information system to keep track of operations were required to complete the census on time and within budget.

Planning and Preparation

The Census Bureau designed the 1990 census keeping in mind the special problems that arise in the census-taking process, as well as constraints of time, budget, and the need to protect individual confidentiality. Plans incorporated the lessons learned from previous censuses. The plans were tailored to implementation and management by a temporary work force in a compressed time frame. Extensive testing was conducted so that hard evidence could be gathered on the utility of new procedures and techniques. The testing also allowed new procedures and techniques to be refined and adjusted.

Formal planning for the 1990 census began in FY 1984. This early start allowed the Bureau to begin major testing of proposed design features earlier for the 1990 census than for the 1980 census (1984 *vs* 1976), and to conduct more major tests of proposed features than for prior censuses (e.g., 7 for 1990 *vs* 5 for 1980). Improvements were made in every phase of census-taking. Some were aimed directly at overcoming operational, control, and timeliness problems identified in 1980 census operations. Others were intended to increase the cooperation of hard-to-enumerate groups. These improvements are described in detail in "Planned Improvements in the Counts for the 1990 Census," April 1989, Bureau of the Census. Improvements included:

- An expanded promotion campaign aimed at hard-to-enumerate groups. For example, for the first time, the Bureau used minority advertising campaigns designed by minority firms, in addition to a more traditional general-audience campaign.
- More cooperation between the Census Bureau and state and local governments. For example, the Census Bureau improved and expanded the Local Review Program, which gives local officials an opportunity to review census counts, by providing training on how to participate in the program, and by instituting two phases of review instead of one, as was the case for the 1980 census.
- Efforts intended to make it easier for people to respond ... census questionnaires. For example, the Bureau expanded questionnaire assistance operations for 1990 by offering toll-free

[126] Ericksen, et al., pages 17–19.
[127] See the discussion in Guideline One above.
[128] See the discussion in Guideline One above, where the deficiencies in distributive accuracy of an adjusted count, using Census Bureau procedures, are detailed.

telephone assistance in English, in Spanish, and in six Asian languages, and by sending out multilingual "early alert" flyers about the census in selected areas.

• Tailoring census procedures to deal with special or unusual situations. For example, enumerators delivered questionnaires to public housing developments, and the Bureau hired public housing residents to deliver the questionnaires and conduct outreach activities at the same time.

• A greatly increased amount of automation in the census. For example, an automated management information system, in conjunction with an automated address control file, enabled home office control and monitoring of the 1990 census to deal with developing problems early and rapidly.

• Implementing an automated geographic control system—called TIGER—in cooperation with the U.S. Geological Survey. The TIGER System solved one of the most serious problems of the 1980 census—late, inconsistent, and illegible maps. The TIGER System assured accurate and timely maps and geographic files for the 1990 census.

The 1988 dress rehearsal was the capstone of planning efforts; it was preceded by 5 years of consultation with data users and formal tests of alternative procedures and questionnaire content of the kind just described. The Bureau consulted with a wide range of data users, including minority organizations, planners and academics, business leaders, representatives of private organizations, state and local officials, and Federal agencies.

Once the basic plan for the census, including improvements, was determined, the Census Bureau began to prepare for 1990 data collection and processing. These preparations included map-making, questionnaire printing, address list construction, setting up a field structure of over 500 offices for data collection and processing, procuring and installing automated equipment, and preparing promotion materials.

A critical activity was preparation of a precensus address list. This list was used to determine which housing units had or had not returned a questionnaire in areas where householders were instructed to return their questionnaires by mail. In all, some 100 million addresses were compiled before the census from purchased lists, field canvassing by census enumerators, and a series of overlapping checks and update operations by census workers, the U.S. Postal Service, and review by local officials.

By March 1990, all preparatory activities had been completed and the data collection phase of the census, which involved attempting to get a completed questionnaire for every person and housing unit in the Nation, was set to begin. (Enumeration of remote areas of Alaska had begun a few weeks earlier in order to complete the enumeration before the Spring thaw.)

Basic Enumeration Procedures

The 1990 census was planned to be a multiphase and incremental process that was to determine the population as of April 1, 1990. Except for remote areas of Alaska, questionnaire delivery or mail-out occurred in March 1990, but the enumeration was not intended to be over then. The Census Bureau built into the census process programs to follow up on housing units that did not return a questionnaire and to ensure that every reasonable effort was made to enumerate every housing unit. These programs extended well after April, into the fall of 1990.

90 percent of the housing units were expected to complete questionnaires and return them by mail. Two procedures were used in such mail-back areas—mail-out/mail-back and update/leave.

For the remaining housing units, householders were instructed to hold their completed questionnaires for enumerator pick-up. This procedure was called list-enumerate. Other special procedures were designed to enumerate persons who lived in group quarters (such as college dormitories and military barracks) and persons who had no usual residence.

Mail-Back Areas

Mail-Out/Mail-Back

The mail-out/mail-back procedure was used for large cities, suburban areas, and some smaller cities, towns, and rural areas where mailing addresses were house number and street name. In all, about 83 percent of U.S. housing units were in mail-out/mail-back areas. Mail carriers in these areas delivered addressed questionnaires on March 23, 1990, and householders were asked to mail back completed questionnaires by April 1, 1990. Five out of six housing units received a short form containing *only* the questions asked of all housing units; one out of six housing units received a long form with additional questions. One week after mail-out, a post card was sent to each housing unit reminding persons to fill out the questionnaire and return it as soon as possible. This was in addition to the multiple-component promotion campaign, then at its peak.

The USPS returned some questionnaires to the Census Bureau as "undeliverable." The Bureau added a special operation to have census enumerators deliver by hand as many of the "undeliverables" as possible. The remaining housing units did not receive a mailing piece at this time, so they were enumerated during nonresponse follow-up (see below).

Update/Leave

The update/leave method was used in rural areas in the South, Midwest, and Appalachia, where mailing addresses are rural-route designations, or where many householders pick up their mail at lock-boxes. These areas contain about 11 percent of the housing units in the Nation. Here, census enumerators, rather than the USPS, delivered the census questionnaires and, at the same time, updated the address list. This operation began in early March 1990 and continued throughout that month. Just as in mail-out/mail-back areas, householders in update/leave areas were to complete and mail back their questionnaires by April 1, 1990. Again, most units received a short form, but a small pre-designated sample received the long form. Householders in these areas also received a reminder postcard asking them to return their questionnaires.

List/Enumerate

The list/enumerate, or door-to-door method, was used for about 6 percent of the Nation's housing units. These units were primarily in very remote and sparsely settled areas. There was no precensus address list for these areas. Mail carriers delivered unaddressed short-form questionnaires on March 23 and, beginning about April 1, census enumerators went door-to-door listing addresses, picking up completed questionnaires or filling out questionnaires as necessary, and administering the long form at a sample of these units.

Special Procedures

Special place enumeration took place in March and April, 1990. Special places include group quarters, such as boarding houses, nursing homes, dormitories, rectories, convents, hospitals, etc. Enumerators visited these places to collect information from each resident. About 2 weeks before Census Day, the Census Bureau also conducted a Street and Shelter enumeration (S-night) to collect information from components of the homeless population. The first phase

of this operation focused on enumerating persons staying in shelters for the homeless, while the second phase focused on enumerating homeless persons living outside of shelters, for example, on the street.

There were two additional components of special place enumeration: Transient enumeration and military enumeration.

- During transient enumeration, census workers visited travel places where guests are unlikely to have been reported at their usual place of residence, or where guests are unlikely to have a permanent residence. These places include YMCA's, YWCA's, youth hostels, commercial campgrounds, etc.
- For military enumeration, special procedures were used to count domestic military and maritime personnel. Military bases and vessels were self-enumerating. In these instances, bases appointed a senior commissioned officer to serve as the enumeration project officer.

Questionnaire Receipt

Some households received a short questionnaire containing only the questions asked of all households, while others received a long form containing additional questions. About 17 percent (or a sampling rate of about 1-in-6) of the households received the long form. However, in places with an estimated 1988 population of less than 2,500, the sampling rate was 1-in-2. Based also on precensus estimates, very populous census blocks had a sampling rate of 1-in-8. All other areas had a sampling rate of 1-in-6.

Once questionnaires had been delivered, forms began to arrive by mail in district or processing offices serving each area. Mail returns for some areas went to a processing office for check-in. For most areas, mail returns, as well as questionnaires completed by enumerators during list/enumeration or special place enumeration, went directly to a district office. Both processing offices and district offices used automated equipment to check in forms by bar code scanning of the return envelope. The associated address in the automated address control file was then coded to show that a questionnaire had been received for that unit. At the conclusion of the check-in phase, each listing not coded represented a case that would have to be visited by an enumerator during nonresponse follow-up.

Nonresponse Follow-up

The Census Bureau followed up every housing unit for which a questionnaire was not returned. Daily reports on the mail return check-in rates for each district office were transmitted to headquarters through the automated management information system. This information was used to project the likely workloads for nonresponse follow-up. This overall workload was expected to require over 250,000 temporary enumerators to visit 30 million units over a 2 month-period. By the end of April, the Census Bureau had to estimate the number of persons it needed to hire, and to begin preparing lists of addresses that had not returned a questionnaire. The mail response rate was 63 percent, lower than the projected 70 percent. As a result of this, the Census Bureau hired more enumerators than it had originally planned for nonresponse follow-up.

The Census Bureau completed nonresponse follow-up for the 1990 census substantially earlier than had been the case for the 1980 census, despite a larger workload. Recruitment goals were met despite the need for more workers engendered by the low mail response rate, and in spite of lower levels of general workforce unemployment than had been the case for the 1980 census.

During nonresponse follow-up, enumerators were required to make up to six attempts to contact a household member and complete a census questionnaire. If this was not possible after three personal visits and three telephone calls at different times and on different days, the enumerator attempted to obtain at least basic information on household member(s) from knowledgeable sources, such as neighbors or building managers.

Because the nonresponse follow-up had to be completed quickly so that other operations could be conducted, each district office was authorized to begin a final phase of nonresponse follow-up once 95 percent or so of the operation had been completed. During this phase, enumerators made one more visit to each remaining case to obtain as complete an interview as possible.

Coverage Improvement Efforts

Basic data collection activities included various steps designed to improve census coverage. Among these were special promotion and outreach efforts, better address listing procedures, extra efforts to increase mail returns, follow-up on all housing units that did not return a questionnaire, better management of and pay for enumerators, etc. But after basic data collection, census plans also included additional special programs to improve the population count that went beyond standard procedures.

These additional coverage improvement programs, which represent the Census Bureau's policy of giving everyone several opportunities to be included in the census counts, added about 5.4 million persons to the census counts, or about 2.2 percent of the total enumerated population.

Such coverage improvement programs included: (1) The 100-percent recheck of vacant housing units or those identified as uninhabitable or nonexistent; (2) the "Were You Counted?" campaign, an opportunity for people who thought they might have been missed to call in or fill out a census form printed in the newspaper; (3) the parolee and probationer check, which involved working with parole and probation officers to get names and Census Day addresses of parolees and probationers and add them to the census had they not already been counted; (4) the housing coverage check, in which the Census Bureau recanvassed selected blocks based on evidence brought to its attention by the automated management information system; and (5) the postcensus phase of the local government review program.

Recheck of Vacant Housing Units and Those Identified as Uninhabitable or Nonexistent

During the follow-up of nonrespondents by enumerators in May through July, some housing units were identified as vacant or uninhabitable; some addresses were added to the address control file. Each of these units was rechecked by another enumerator in July or August.

Of the approximate 8 million vacancies, the recheck showed 7.6 percent had been occupied as of Census Day, April 1. Their occupants were enumerated at the time of the recheck. This added about 1.6 million persons to the count. Of the approximate 2.9 million units previously identified as uninhabitable or nonexistent, 5.4 percent were reinstated as occupied April 1. These conversions added almost one-half million persons to the count.

"Were You Counted?" Campaign

After the primary data collection, the Census Bureau initiated a procedure to give anyone who thought he/she had been missed the opportunity to fill out publicly available forms or call toll-free 800 numbers that operated in English, Spanish, and six Asian languages. Communities, the media, and many of the 56,000 community-based organizations that had helped initially promote answering the census were encouraged to conduct "Were You

Counted?" campaigns, reproduce census-designed forms or promote call-ins to the 800 numbers. The purpose of the campaign was to give a second chance to those who might initially have avoided being counted, or to reach persons not part of the principal family in a household who might not have been listed on the household questionnaire. Initially, the Census Bureau planned to end the campaign by June 30, 1990, but because so many organizations participated, the toll-free numbers were held open until September 30.

In all, about 400,000 "Were You Counted?" calls or forms came into the Census Bureau. Although the majority of these proved to be persons who had already been counted, the forms did add over 200,000 persons to the census.

Parolee and Probationer Count Check

Research had suggested that a group with a high probability of having been missed in prior censuses were those on parole or probation, a group consisting disproportionately of young males. Thus, in February 1990 the Census Bureau sent letters to the governors and heads of correction departments in each state and the District of Columbia asking them to participate in a program to get parolees and probationers counted. Each was asked to name a liaison to handle the program. Each liaison was sent special individual forms to distribute to their parole and probation officers, who in turn were to distribute them to those under their jurisdiction.

The response rate for the program was disappointingly low—so low in fact, that the Census Bureau sent enumerators to work with parole and probation officers to complete a form for each parolee/probationer with a verified April 1 address. As a result of this activity, it is estimated over 400,000 persons were added to the census.

Housing Coverage Check

With a computerized census that captured questionnaire data as returns came in, it was possible to make additional accuracy checks not possible in prior censuses. In August of 1990, the Census Bureau searched its data bases to identify any blocks or communities for indications of a low count. While the census was still in progress there was time for a further canvass to make corrections. Population and housing counts, which had accrued thus far for the 39,189 units of local governments, were compared with 1980 counts and recent population estimates. The Census Bureau looked at its data on areas of new construction for possible missed new subdivisions. It also searched to see if the "Were You Counted?" forms showed any pockets of housing that might have been missed. It looked at media reports or local complaints of missed buildings or blocks. Based on these data searches, the Census Bureau decided to recanvass blocks where problems might exist. These blocks represented 15 percent of the Nation's housing units.

Postcensus Local Government Review

39,189 units of local government were sent housing counts and group quarters counts, accrued as of mid-August, to compare with local data. (New updated maps for the communities had already been sent to them in July). Governments were given 15 working days in which to challenge the housing unit or group quarters count for any block. The feedback from local governments was varied. Many took the counts to be final, although the Vacancy Recheck, the Housing Coverage Check—in fact all of the coverage improvement projects done after the primary data collection—were still in progress. All in all, 17 percent of local governments, including all of the 51 largest cities, challenged some blocks, and eight cities challenged over 2,000 blocks. Cities that challenged more than 2,000 blocks in Postcensus Local Review were Atlanta, Boston, Chicago, Detroit, Honolulu, Los Angeles, New York, and Philadelphia.

The recanvass generated by the Housing Coverage Check and Local Government Review yielded new housing units that added over 300,000 persons to the final census count.

The 1990 Post-Enumeration Survey (PES)

Background

The Census Bureau used two major programs to measure coverage for the 1990 census. The first was the Post-Enumeration Survey (PES), which was an independent survey taken after the census and then compared to the census to attempt to measure coverage error in the census. The second program was Demographic Analysis (DA). DA produced an independent estimate of total population by combining information from various sources of administrative data. The process included using historical data on births, deaths, and legal immigration combined with estimates of emigration, undocumented immigration, and Medicare information. Estimates of total population from DA were then compared with census counts to get an estimate of coverage error.

Summary

The PES was a check of the census but not a recount. After the census, interviewers returned to the field to identify all persons living in the sample of blocks at the time of the PES. During the interview, the interviewer asked where each person was living on Census Day—April 1, 1990. This information was then matched to actual census questionnaires. Most people on the PES questionnaires matched to the census. Some did not, and these are the people estimated to have been missed in the actual census. This part of the PES was called the P-sample. People estimated to be missed based on the P-sample were estimated gross omissions in the census.

People can also be included in the census erroneously. An erroneous census enumeration, for example, could be a child born after April 1, 1990, a person who died before April 1, or a college student away from home who was enumerated at his or her parents' address instead of being correctly enumerated at his or her college. Erroneous enumerations also include persons counted twice in the census. Gross erroneous inclusions in the census were measured in the same blocks as the PES and were called the E-sample.

The data on gross erroneous inclusions and gross erroneous omissions were used to produce an estimate of the net undercount or net overcount of the population in the census. This process is described in the following paragraphs. [1]

Selecting the Sample (Sample Design)

The census attempted to cover all people and was conducted in all blocks. The PES was a sample. The PES sample was selected in stages. First a random sample of blocks was chosen. Within sample blocks, all housing units were interviewed. Within an interviewed housing unit, a PES interview was conducted for each person.

Since the PES was a sample, if total population estimates were to be calculated based on it, the results had to be generalized to other people not living in sample blocks. One statistical method to improve the accuracy of this generalization process was to classify sample cases into groups (called post-strata) such that within a group, people were as alike as possible with regard to their propensity to be undercounted. Ancillary evidence indicates that undercoverage is worse for males than

[1] For a more detailed discussion of PES see Howard Hogan, "The 1990 Post-Enumeration Survey: An Overview," a paper presented at the American Statistical Association in August 1990.

females; for minorities than non-minorities; for renters than owners, etc. Therefore, these types of characteristics were used to define the post-strata. The Bureau did not know which post-stratum to assign a person to until after the PES interview was conducted. To help insure an appropriate sample size by post-stratum, the blocks in the U.S. were stratified by similar characteristics before selecting the sample blocks from them.

All blocks in the United States were assigned to one of 101 strata. The strata were defined by geography, city size, racial composition, and percent renter. A representative set of blocks was selected from each stratum. A separate sampling stratum was defined for American Indian Reservations.

Persons living in institutions were excluded from the PES, as were military personnel living in barracks, people living in remote rural Alaska, persons in emergency shelters and persons who had no formal shelter. For each of these categories, it was unreasonable to expect to be able to conduct an independent interview in July and match them to their April 1 location.

The eventual PES sample consisted of about 168,794 housing units in 5,290 block clusters that included 12,124 blocks. (See attachment 1, "PES Sample Size by State.")

The sample was designed to achieve a .7 percent coefficient of variation. That is, the level of sampling error was expected to be .7 percent of the level of estimated undercount or overcount. So for example, if the PES estimated the undercount to be 5 percent, it was expected that the sampling error (or margin of error) on that estimate would be .35 percent. In practice, the sampling error was, on average, 1.7 times more than anticipated by the sample design.

Listing and Enumerating

In February 1990, permanent interviewers of the Census Bureau visited each of the sample blocks to list all housing units they contained. To preserve independence, none of the temporary enumerators hired to take the 1990 census was used for this operation; nor was the listing conducted out of the temporary census offices. To maintain independence, the Census Bureau did not want anyone to know where a PES sample block was so that it would be treated differently during the census.

After the completion of the 1990 census follow-up of those housing units that did not return a questionnaire (called nonresponse follow-up), a set of PES enumerators interviewed persons at households in the PES sample blocks. Although this interviewing drew from enumerators who had worked on 1990 census follow-up, steps were taken to preserve independence, such as not allowing an enumerator to work in a block in the PES that he or she had worked in during the census.

The interviewers determined who was living in each housing unit, obtained their characteristics, and asked where they lived on April 1, 1990, Census Day. The PES interviewing began nearly 3 months after Census Day. Many people had moved during that time. In order to determine whether they were enumerated in the census, the Bureau needed to know where they lived on Census Day and, thus, enumerators asked a series of probing questions to determine occupants' Census Day addresses.

There was a quality assurance program for the interviewing phase to ensure that the interviewers really visited the household and that the people listed were indeed real. If interviewers made up people, they would not match to the census and would inflate the undercount rate.

Matching

The next step was to match the persons enumerated during the PES (the P-sample) to the census. The matching operation was the first step in determining whether persons in the P-sample were enumerated by the census or missed. Basically those persons in the P-sample matched to the census were considered to have been enumerated; those nonmatched were considered to have been missed.

Matching was carried out in four stages. It involved an initial stage of computer matching followed by two stages of clerical matching to attempt to resolve cases that the computer could not match. The two stages of clerical matching were differentiated by the level of skill and judgment required to establish a match.

Those persons in the P-sample not matched to the census by computer and the first two stages of clerical matching were assigned for a follow-up interview, if it was determined that additional information was necessary to establish whether a match to the census was appropriate. An additional fourth stage of clerical matching was then conducted that allowed the more skilled clerical matchers to use the information from the follow-up interview to establish additional matches.

First, the matching classified people as included in the census only if they were counted at the address where they should have been counted, according to the information they provided. This concept was called "correct address" matching. For example, census rules required that a college student be enumerated at the university dormitory, not at his/her parents' home. The PES counted the student as "enumerated" only if he/she was counted at the university. If he/she was not counted at the university, then the student was classified as "omitted" even if he/she were counted at home. In order for the estimation to work out, the enumeration at home was classified as erroneous and subtracted from the census. So in this example, there would have been one omission (at the university) and one erroneous enumeration (at home). The two netted out in the aggregate. The decision to use "correct address" matching was not lightly taken. Indeed, some earlier tests used "any address" matching, i.e., attempting to search all reported addresses. Either approach has advantages and disadvantages.

The second concept was that of the search area. If a person reported that he lived at a given address, then the matching classified him as correctly enumerated if he was counted anywhere in the block. It also classified him as correctly enumerated if he was counted in a surrounding block. There was a limit to how far the matching process could search. If a census computer operation coded the address across town, for example NW vs. SE, the matching did not search there and did not find the person. The matching counted him/her as missed. To balance, the system had to count the other enumeration as erroneous, because it was outside the defined search area.

A final concept was the idea of "sufficient information for matching." When a match was found, it was easy to say that the case was enumerated. When no match was found, it did not necessarily prove that the person was not enumerated, but merely that the search had not been conducted in the correct place. A further review of the case might have shown that there was "insufficient information," leading to its being classified as "unresolved." Rules that classify cases as "sufficient information for matching" were applied before the matching begins. These rules were designed so that for matches there was confidence that the person was correctly enumerated and, equally important, for non-matches, there was confidence that the person was omitted. This approach leads to a somewhat higher "unresolved" rate, but presumably to more accurate overall results.

The accuracy and consistency of the matching process were central to the PES process. Too many matches would

have decreased the estimate of population, too few would have increased it. Matching errors would have distorted the estimated population distribution if they differed by post-strata. The rules were developed over a decade of research. The multiple levels of matching were designed to ensure that the rules were applied consistently between clerks and between offices.

The E-sample, those persons in the PES blocks who were enumerated in the census, was examined to determine if they were correctly enumerated. E-sample persons were matched back into the census to determine if they were enumerated more than once (duplicates). E-sample persons who were matched to the P-sample were assumed to be correctly enumerated (except for duplicate census enumerations). The remaining E-sample persons who were not matched to the P-sample were potential candidates for erroneous enumerations. These unmatched census persons were also included in the PES follow-up operation described above. The follow-up interviewers determined the enumeration status of those persons; that is, if they were correctly enumerated and simply not in the P-sample or if they were erroneously enumerated.

Errors in measuring census erroneous enumerations have almost as much effect on the final estimate of net undercount as errors in measuring census omissions. Reinterview and rematch studies were used to measure the error that the PES makes in measuring census erroneous enumerations and the effects of these errors on the PES estimates.

In processing the E-sample, it was important to include all census enumerations, especially those conducted long after April 1. Common sense and the results from 1980 both indicated that these were more likely to be erroneous than those done on or near April 1. Because of this, there was a special operation to process census enumerations that were enumerated late in the census process. This operation presented special challenges in merging the data with the results of the earlier operation and completing the processing in time.

A final matching and reconciliation operation took place at the conclusion of the PES follow-up. This included the fourth stage of clerical matching for the P-sample and a determination of whether persons in the E-sample were correctly or erroneously enumerated. An important aspect of this operation was that situations arose where correct match status for persons in the P-sample, or correct enumeration status for persons in the E-sample, could not be determined. This situation occurred because the initial interview was inconclusive or because an incomplete interview was obtained during the follow-up.

Imputation and Dual System Estimation

A final PES file was created that reflected the results of the operations described above. This file included the characteristics of each person in the P-sample and the E-sample. The file also included the match status for persons in the P-sample and the enumeration status (correct or erroneous) for persons in the E-sample. As the final file was prepared, computer editing or imputation was performed to correct, insofar as possible, for missing or contradictory data. A critical aspect of imputation involved the estimation of a final match status for those persons whose match status could not otherwise be resolved. The estimation of match status was very critical. For example, mistakes in the PES matching process, which incorrectly identified persons as not counted in the census (nonmatches), erroneously overstated the estimated undercount and vice versa.

The data in the final PES file were then summarized and incorporated with data from the full census to produce dual system (PES and census) estimates (DSE's) of total population. The DSE's were produced for unique estimation strata (or groupings of persons described below). The dual system estimator is explained more fully in Hogan's document cited above. Essentially it involves estimating how many people were (1) in the PES and in the census, (2) in the PES and out of the census, (3) in the census but not in the PES, and (4) in neither the census nor PES.

The dual system model conceptualized each person as either in or not in the census enumeration, as well as either in or not in the PES. Each person was classified according to the following tableau where the subscripts denote row and column and the stars indicate summing over the entire row/column. $N_{\bullet\bullet}$ denotes the entire population.

ENUMERATION

PES	Total	In	Out
Total	$N_{\bullet\bullet}$	$N_{\bullet 1}$	$N_{\bullet 2}$
In	$N_{1\bullet}$	N_{11}	N_{12}
Out	$N_{2\bullet}$	N_{21}	N_{22}

All cells were conceptually observable except for N_{22}, and of course any of the marginal totals that include N_{22}. The cell N_{22} (often called the 4th cell) was an estimate of people missed in both the census and the PES. Even though not directly observable, the DSE of total population included an estimate of people in the 4th cell. The DSE of total population was based on several assumptions. If the PES was an (approximately) unbiased sample of the whole population, then an (approximately) unbiased estimate of $N_{\bullet\bullet}$ could be made by noting that the ratio of those in the PES and in the census to the total in the PES should have been the same as the ratio of the total in the census to the total population. Algebraically:

$$N_{11}/N_{1\bullet}=N_{\bullet 1}/N_{\bullet\bullet}$$

Then solve for the total population:

$$N_{\bullet\bullet}=(N_{\bullet 1}\cdot N_{1\bullet})/N_{11}$$

This is the dual system estimator of total population.

DSE's were prepared in each of 1,392 post-strata (see next section for a description). Knowing the undercount or overcount rate for each of the groups was important for estimating the net undercount at the local level. It was acceptable for both the PES and the census to have different coverage rates for different post-strata. However, if within a post-stratum, there were subgroups where both the PES and the census had significantly lower coverage, then the DSE would have been biased.

Another type of bias would have arisen if being enumerated in the census affected the person's response to the PES, or being in the PES affected the person's response to the census enumeration. This would be the case if the PES interviewer and the enumerator compared notes, or if a person refused to cooperate in the census because he had been recently interviewed in PES. The design sought to minimize this effect by conducting the PES after most of the census operations were completed and by conducting the PES out of the Regional Census Centers rather than out of the local District Offices that conducted the enumeration.

Post-Strata

Using the match status and key data, such as age, race, and sex for each person in the sample, the Bureau prepared DSE's of the total population for each of 1,392 groupings of people (post-strata). The reason for forming the post-strata was to group persons who had similar chances of being enumerated in the census. The post-strata were defined by census division, geographic subdivisions such as central

cities of large metropolitan statistical areas, whether the person was the owner or renter of the housing unit, race, age, and sex. Each person in the PES sample belonged in one of the unique post-strata. A full description of the 1,392 post-strata is shown in attachment 2.

For purposes of illustration, the following are examples of the 1,392 post-strata. One example is a post-stratum which contains Black males, age 20–29, living in rented housing in central cities in the New York primary metropolitan statistical area. A second example is that which contains non-Black non-Hispanic females, age 45–64, living in owned or rented housing in a non-metropolitan place of 10,000 or more population in the Mountain Division. A third example is that which contains Asian males, age 45–64, living in owned or rented housing in metropolitan statistical areas but not in a central city in the Pacific Division. A fourth example is that which contains non-black Hispanic females, age 30–44, living in owned or rented housing in central cities in the Los Angeles-Long Beach primary metropolitan statistical area or other central cities in metropolitan statistical areas in the Pacific Region. As can be seen from these examples, the 1,392 post-strata are very specific.

The Decision on Combining PES and DA Results Before Computing Adjustment Factors

It was expected that the estimate of total population from the PES would be lower than the estimate of total population from DA. That is because there is a tendency for some people to be missed in both the census and the PES. (often referred to as correlation bias.) No such bias exists with DA estimates. For that reason, there was an open decision point about whether or not to "rake" PES estimates to DA estimates before producing adjustment factors.

After examining the information, the Census Bureau decided against trying to combine the results of DA and PES. There were several reasons for the decision. Some of the main ones include:

- The PES estimate of total population was higher than the DA estimate.
- The PES estimate of females was considerably higher than the DA estimate.
- At the point in time the decision had to be made, the DA estimates were preliminary. There was concern that DA estimates might change considerably over time.
- A concern about the quality of certain components of the DA estimates; for example, the estimate of undocumented immigrants.
- The uncertainty about how combining DA estimates might effect the assumptions underlying the DSE system.

Adjustment Factors

The next step in the post-enumeration survey process was to compare the estimated total population for each post-stratum (the dual system estimate or DSE) to the census count to determine a "raw" adjustment factor. For example, if the DSE for a particular post-stratum was 1,050,000 and the census count was 1,000,000, then the adjustment factor was 1.05, reflecting about a 5-percent estimated net undercount of variability. An adjustment factor may be less than one, thus lowering the census count in a post-stratum if an adjustment is applied. This results when there is evidence of an overcount in the post-stratum.

"Smoothing" the Adjustment Factors

The next steps were "smoothing" the variances of these "raw" adjustment factors, "smoothing" the "raw" adjustment factors themselves to reduce sampling variance associated with them, and the production of final adjustment factors incorporating both smoothing steps. Because the PES was a sample, it was subject to sampling error. Sampling error is an estimate of the error associated with taking some of the population (a sample) rather than all of the population (a census). Disaggregating 377,000 PES persons to 1,392 post-strata produced some post-strata with small sample sizes, and therefore, high estimates of sampling error. The process of smoothing the "raw" adjustment factors to create final adjustment factors was a step to minimize the effect of sampling error.

Both "smoothing" steps were based on a multi-variate regression model. The factor smoothing step used observed characteristics that have been known to be correlated with undercount. A regression prediction model "predicted" the adjustment factor for each of the 1,392 post-strata. The final adjustment factor was then a weighted average of the originally observed adjustment factor (called "raw") and the modeled factor (from the regression prediction model.) For a post-stratum with low estimated sampling variance, there was heavy weight on the observed factor; and vice versa. The final adjustment factors by post-stratum are shown in attachment 3.

Small Area Estimation

The final adjustment factors were now ready to be used to produce adjusted counts for every block in the Nation. The PES can only produce "direct" estimates of the total population for relatively large geographic areas (i.e., the 1,392 post-strata). If there is a decision to adjust, however, the adjustment must be applied to each of the Nation's 4 million populated blocks. The Bureau developed a model that takes the adjustment factors produced for each of the 1,392 post-strata areas and uses them to estimate adjustment counts for each block. Since each of the post-strata crosses many blocks, the Bureau based its model on a critical assumption that coverage error is similar for all blocks that a post-stratum crosses.

Here are two examples of how block counts could be changed during this process. Suppose a census block with 200 people had 50 people who fell into a particular post-stratum. An adjustment factor of 1.05 was computed for that post-stratum, so 50 was multiplied 1.05, which comes to 52.5. Since procedures allowed adding only whole persons to a block, either 2 or 3 persons were added, based on a pre-specified procedure, to the persons in that post-stratum for that block. Other groupings of persons in the block in this example also were multiplied by the adjustment factor for the post-stratum into which they fell. Similarly, suppose there were 80 people in another post-stratum in a particular census block, and the adjustment factor was 0.94, indicating an overcount. 80 was multiplied by 0.94, which came to 75.2, so 4 or 5 person records were eliminated from that block.

The Bureau then produced a data file that included enumerated people plus people added (or subtracted) by adjustment. It did this by adding or subtracting "adjustment" persons with characteristics that were imputed from other persons in the same block. The "adjusted" data files could then be used to produce all required census tabulations.

The 1990 Post Enumeration Survey Evaluation Program

The Post Enumeration Survey (PES) was conducted to evaluate the coverage of the 1990 Decennial Census. Twenty evaluation projects were subsequently conducted to evaluate the PES.[2] This report briefly describes the objectives and implementation of these twenty PES evaluation projects.

[2] In this document, studies P-13 and P-14 are discussed as one study each, although each had two parts. Elsewhere, these parts may be discussed separately, which leads to a total of twenty-two studies.

Ten of the sources of potential error in the PES were addressed by the evaluation studies:

1. Missing Data.
2. Quality of the Reported Census Day Address.
3. Fabrication in the P-sample.
4. Matching Error.
5. Measurement of Erroneous Enumerations.
6. Balancing the Estimates of Gross Overcount and Gross Undercount.
7. Correlation Bias.
8. Small Area Estimation.
9. Late Census Data.
10. Total Error.

Each of these ten potential sources of error are herein described along with the specific PES Evaluation project used to evaluate or estimate that error.

More detailed project descriptions are found in the Project Plans dated July 31, 1990. For more detailed descriptions of the implementation and results of these projects, see the final reports of July, 1991, whose executive summaries can be found in Appendix 3.

1. Missing Data

Both the P- and E-samples contain missing data on enumeration status. The E-sample has cases where the information required to determine whether the person is correctly or erroneously enumerated in the census is not available. The P-sample has cases where the information needed to determine whether the person is enumerated in the census is not available.

Missing data occur in more than one way. The interviewer may be unable to obtain an interview during the P-sample interview or during the PES follow-up. A P- or E-sample questionnaire may not have all the demographic and housing information to establish correct enumeration status. Finally, even with all the information requested on the questionnaires, circumstances may be so unclear that the enumeration status cannot be resolved or determined.

Missing data on enumeration status were handled in the production PES in three ways: noninterviews to the P-sample interview were handled by a weight adjustment; missing demographic characteristics in the P- and E-samples (such as age or race) were imputed by means of a hot-deck procedure; and unresolved match status cases were handled by a logistic regression technique.

Missing data can affect the estimates of undercount in a number of ways. For example, if the number of imputed correct enumerations is too high, the undercount estimate will be biased upward, or if the number of imputed matches in the P-sample is too high, the undercount estimate is biased downward.

Project P1: Analysis of Reasonable Alternatives

The analysis was based on applying alternative missing data treatments, such as methods of handling proxy interviews and mover data, applying bootstrap samples and applying other logistic regression methodologies to study the sensitivity of the dual system estimate to the method of imputation of missing data. A narrow range of alternative estimates indicates robustness in the dual system estimates, indicating little uncertainty in the estimates due to missing data.

The following were the principal alternate imputation treatments:

P-sample Proxy Alternative: P-sample follow-up interviews marked as proxies (i.e. completed with nonhousehold member) were recoded to indicate that no interview was obtained during follow-up.

E-sample Proxy Alternative: E-sample follow-up interviews marked as proxies (i.e. completed with nonhousehold member) were recoded to indicate that no interview was obtained during follow-up.

P-sample Mover Alternative: Unresolved P-sample movers were imputed as if they were nonmovers.

1988 Style Logistic Regression Alternative: The 1990 production imputation model is quite different than the model that was used in the 1988 Dress Rehearsal. The 1988 Style Logistic Regression Model consists of several standard logistic regression models as in 1988.

Bootstrap Samples: Three E-sample and three P-sample bootstrap samples were drawn in order to measure the variation in the production dual system estimates given the PES sample of blocks. Each bootstrap consisted of selecting households with replacement within blocks.

Imputation Treatment Combinations: Dual system estimates were computed for imputation treatment combinations. The following treatment combinations were used:

P-sample Proxy and E-sample Proxy
P-sample Proxy and 1988 Style Model
E-sample Proxy and 1988 Style Model
P-sample Proxy, E-sample Proxy, and 1988 Style Model

Project P2: Distribution of Missing Data Rates

This study was based on analysis of the missing data rates observed for the P- and E- samples. The types of missing data of greatest interest are noninterviews for the initial PES interview, and unresolved cases which remain after the PES follow-up.

The objectives of PES evaluation project P2 are to determine the level and distribution of missing data by demographic and geographic breaks and to compare the distributions with the distribution of census undercount (overcount). Hence, the following estimates are examined for P2.

1. Outcome of Interview (PES, PES Follow-up, and PES Evaluations).
2. Proxy Rates (PES, PES Follow-up, and PES Evaluations).
3. Percentage of Item Imputation (Hot-Deck and Logistic Regression).
4. Correlation Between Item Imputation and Census Undercount.

Project P3: Evaluation of Imputation Methodology for Unresolved Match Status Cases

This study was based on a reinterview of a sample of the P- and E-sample cases that were unresolved after the completion of the PES production follow-up. The reinterview also included a sample of the initial PES incomplete interviews. The reinterview was conducted immediately following the final PES matching operation. The reinterview used a probing questionnaire and better quality interviewers. In addition, the reinterview procedure allowed greater opportunity to contact knowledgeable respondents.

The objectives of PES evaluation project P3 are to: (1) provide quantitative information on the effect of the match/enumeration status imputation procedures; (2) examine quantitative measures of the effect of the noninterview adjustment; and (3) examine the characteristics of the household noninterviews. Hence, the following aspects of the PES are evaluated in P3.

1. Match/Enumeration Status Imputation.
2. Converted PES Noninterview Households.
3. PES Noninterview Household Characteristics.

2. Quality of the Reported Census Day Address

Dual system estimation assumes that P-sample respondents can be linked, or matched, correctly to their census day address. This evaluation measures address reporting and the error in the number of people matching a census enumeration due to address reporting error. Census Day was on April 1, 1990. The PES was conducted in July and August, 1990. Thus, some of the

respondents had moved between the time the census was conducted and the PES was in the field. However, in spite of probes on the PES interview questionnaire, respondents may fail to report that they moved. This type of error may cause the matching operation to search the census in an area other than where the respondent was enumerated and to assign a nonmatch status to respondents who might have been enumerated.

Project P4: Quality of the Reported Census Day Address—Evaluation Follow-up

An additional reinterview of a sample of P-Sample cases from the production follow-up was conducted. The sample consisted of nonmatches and unresolved P-sample cases in the PES block clusters selected for the evaluation follow-up. Some matches from whole household matched households were subsampled within each cluster. In addition, matches were selected from partially matched households. A specially designed questionnaire with special probes was used by highly skilled enumerators (Census Bureau Field Representatives). The reinterview allowed greater opportunity to contact designated respondents and probe more deeply for census day accuracy of the PES process for identifying movers and the quality of mover address reporting. Therefore, reviewing these results allowed an assessment of the accuracy of the census day address reported in the production PES.

This evaluation is based on a follow-up and reinterview operation that took place immediately following the final PES matching operation. The follow-up operation consisted of a sample of P-sample matched and nonmatched persons who were excluded from the production follow-up. A review of the results of this follow-up addressed the questions concerning the assumptions underlying the rules that were used in determining which cases should be sent for the production follow-up. This operation was done after PES production matching had been concluded.

3. Fabrication in the P-Sample

Interviewers, for whatever reason, may fabricate persons within enumerated housing units. The PES program had an extensive quality control (QC) program that identified and corrected fabrications. However, even with the best of intentions fabrications potentially remain after this operation. Three studies were implemented to address the effect of any uncorrected fabrications that remained in the data set after the quality control operation. The first study (P5a) identifies the residual fabrication by means of the evaluation follow-up and revisit interviews; subsequent matching of these households will identify fabrications. The second study (P5) utilizes the PES field operation quality control records to estimate "upper bound" residual PES fabrications. The third study (P6) provides model-based estimates of fabrications by comparing, at the block level, interviewer nonmatch rates with "nearby" interviewer nonmatch rates. These comparisons provide an indication of the quality of the interviewers work.

Project P5a: Analysis of P-Sample Fabrication From Evaluation Follow-up Data

The evaluation follow-up described for Project P–4, provided estimates of P-sample fabricated persons. These estimated fabrications can be used as independent estimates (from the quality control) of the level of fabrications in the P-sample. In addition, the quality control operations for the PES interviewing were assessed by comparing the estimated residual error rate from quality control records with the estimated fabrication rate from the follow-up.

Project P5: Analysis of PES P-Sample Fabrications From PES Quality Control Data

The data for project P5 comes from the Quality Control operation of the PES interviewing phase. The purpose of the QC check is to confirm that the PES interviewer visited the correct housing unit and conducted the interview according to the survey procedures. The roster of names, ages and census day addresses are all verified during the interview for the QC sample. A P-sample questionnaire fails the QC check when the household roster is incorrect. When an error is detected, all the recent work of the production interviewer undergoes a QC reinterview. Fabricated households discovered as a result of the QC reinterview are not used and correct interviews are obtained. Overall, approximately 35 percent of the P-sample (i.e., 56,000 households) were reinterviewed in the QC operation of the PES interviewing phase through telephone calls and personal visits.

The central problem or assumption of investigation for project P5 is the estimation of the amount of residual (i.e., undetected) fabrication that exists in the P-sample after the QC operation has been concluded. This analysis provides estimates both in terms of households and persons within these households.

Project P6: Fabrication in the P-Sample: Interviewer Effect

The objective of P6 was to gain knowledge about possible undetected fabrication in the PES. Though it is expected that curbstoners make up only a fraction of the PES work force and the quality control detects and eliminates such curbstoning, the potential impact of undetected fabricated data can be serious. This type of error inflates the undercount estimate. In addition, the inflated nonmatch rates are likely differential, i.e., larger for some post-strata than others.

The purpose of this study was to evaluate the quality control procedure implemented in PES to see how effective it was in detecting fabrication. This was done by developing a model to predict the nonmatch rate from the actual nonmatch rate obtained by interviewers working in areas with households of similar demographic characteristics. The assumption underlying the model was the interviewers working in similar areas would have similar nonmatch rates and the deviations from the model would indicate undetected curbstoning. Standardized scores (Z-scores) were computed for each interviewer rather than comparing the absolute differences between the observed and the expected rates. This was done to take into account the size of an interviewer's assignment. Interviewers with large scores differed greatly from the model predication, and were identified as potential curbstoners or poor quality workers. These enumerators were further studied to determine where they had worked and whether they had been detected by the PES QC operation.

4. Matching Error

Errors can occur in the operation where P-sample persons are matched to the original census enumerations. This matching operation was conducted in seven processing offices (PO's). Even though great efforts were made to standardize this operation across all PO's, errors could be relatively concentrated. Two studies were conducted to examine this type of error. The first study (P7) utilized a team of professionals to dependently rematch a subsample of PES block clusters; this operation is referred to as the Matching Error Study. The rematchers had access to the match codes assigned by the PES production matchers, and worked on assignments in PO's other than their home PO where they worked on PES production. The rematch was designed

to estimate the net error rate in the assignment of enumeration status in the P-sample and the E-sample. The second study (P5) examined PES production quality control records. This analysis provides insight into the nature of PES production matching error by examining where differences occur within this multi-tiered operation.

Project P7: Estimates of Clerical Matching Error From the Evaluation

This evaluation was based on a rematch of a subsample of the PES blocks by highly skilled personnel. This project also allowed additional field work as required, when additional information was determined to be necessary to resolve specific cases. The assumption underlying the evaluation is that better training and personnel can detect systematic errors in the matching.

The subsample of blocks included in this evaluation was based on a stratified sample designed to give a higher probability of selection to blocks with potential matching problems. In addition, the highly skilled personnel used for this evaluation were assigned to work in different processing offices, to the extent possible, to minimize redoing blocks that they previously processed.

Project P8: Matching Error—Estimates of Clerical Matching Error in the P-Sample From Quality Assurance Results

This evaluation was carried out by comparing the results of the PES matching quality control operation to determine where potential inconsistencies existed.

At the conclusion of the computer matching, the clerical matching proceeds with an initial stage of clerical matching (CMG) followed by a more extensive stage of matching by another group of more qualified special matching group clerks (SMG1). Another special matching group (SMG2) also conducted matching on the same cases as the CMG and SMG1 stages. Discrepancies between the SMG1 and SMG2 are adjudicated by a higher level PES matching technician.

Comparing the differences between the various stages of matching can identify potential areas where matching error can exist. These findings may be of interest in interpreting the results of project P–7.

5. Measurement of Erroneous Enumerations

Some census enumerations are in fact erroneous. The following enumerations are erroneous:

(1) Duplicated persons.

(2) Fictitious persons.

(3) People who died before Census Day.

(4) People who were born after Census Day.

(5) People enumerated outside the search area where they were living on Census Day.

An estimate of erroneous enumerations is needed for the PES-census dual system estimate of the total population. Three studies investigate errors in classifying the enumeration status (correct or erroneous) of the E-sample persons. The first study (P10) utilized the same team of highly skilled professionals as did project P7 to dependently review the PES E-sample production results in a subsample of PES block clusters. This operation was part of the Matching Error Study. The focus was on the errors that occurred during PES production processing involving duplicates and fictitious persons; however, there was also an examination for the above (3), (4), and (5) type errors. The second study (P9a) utilized data collected from the evaluation follow-up interviews. The evaluation follow-up questionnaire was administered by more competent interviewers than was used by PES production. Also, this questionnaire had more probes than the standard PES production follow-up questionnaire. An alternative estimate of erroneous enumerations resulted from this operation. The third study (P9) is a consistency check; an examination of PES E-sample cross-tabulations provides evidence as to whether a particular type of error in classifying enumeration status is present in the data.

Project P10: Accurate Measurement of Census Erroneous Enumerations—Clerical Error in Assignment of Census Enumeration Status

This evaluation was conducted as part of the rematch work described for Project P7, Evaluation of Clerical Error in the P-sample matching. The study used the same subsample of PES blocks. The E-sample for these blocks underwent the intensive review by highly skilled matchers. This work was supplemented by the reinterview described for Project P9a. The objective was to determine whether the production matching operations are correctly classifying census erroneous enumerations.

The combination of both of these projects—P7 and P10- is referred to as the Matching Error Study (MES).

Project P9a: Accurate Measurement of Census Erroneous Enumerations—Evaluation Follow-up

A sample of E-sample cases was sent for a PES evaluation field follow-up to determine whether a person was correctly enumerated in the Census. The sample included both E-sample cases where an interview was obtained and those where a follow-up interview was not completed. The follow-up reinterview was conducted with more experienced enumerators using a more probing questionnaire. In addition, the follow-up allows greater opportunity to contact a respondent and obtain a complete interview. This same evaluation follow-up was used as part of Project P7 and Project P4. The completed evaluation follow-up interview was clerically matched back to the census to assess the accuracy of the PES production procedure in classifying a persons enumeration status.

Project P9: Accurate Measurement of Census Erroneous Enumeration—Consistency Checks

This evaluation was based on examining a variety of cross tabulations prepared from the PES E-sample for each evaluation stratum. Data such as the following was cross-tabulated:

(1) Enumeration status (correct enumeration, erroneous enumeration).

(2) Type of respondent (original census residents, current residents, neighbors, other proxies).

(3) Source of census enumeration (mailback, enumerator return).

(4) Age group.

(5) Enumeration status of other household members (whole household erroneously enumerated, partial household erroneously enumerated).

The cross tabulations were examined to assess whether the pattern of erroneous enumerations was consistent with previous experience and research findings. Unexplainable discrepancies in the erroneous enumerations were considered as potential indications that the PES process incorrectly measured erroneous enumerations.

6. Balancing the Estimates of Gross Overcount and Gross Undercount

Because of the limited search area that is used to estimate P-sample nonmatches and E-sample erroneous enumerations, balancing error can occur. There was no plan to obtain a direct estimate of this type of error. The components of balancing error are included in the measures of errors that are produced from other studies such as P–7 and P–10 (matching error studies)

Project 11: Balancing Error Evaluation—Percentage of Matches Found Outside Sample Blocks

This evaluation used supplementary information to assess whether balancing is an issue in the performance of PES. Inconsistencies found are indications of potential failure of balancing and should be indications of which of the evaluation studies should reflect these errors. The P-sample match rates for the PES blocks and surrounding blocks were compared with the rates at which E-sample persons are found to be in the PES blocks and in the surrounding blocks. These rates should be the about the same. Differences found were evaluated using the results of the evaluation follow-up.

The rate at which movers matched in the blocks to which they were geocoded was also studied. These rates should be consistent with the corresponding rates for the P-sample nonmovers in the same post-strata.

7. Correlation Bias

The dual system estimation used for the PES is based on several independence assumptions. Two that are of particular interest are homogeneity and causality. The homogeneity assumption requires that everyone has the same probability of inclusion in both the P-sample and the census within the same post-stratum. Failure of the homogeneity assumption usually is seen in an understatement of the undercount for a population group (such as Black males). The causality assumption requires that inclusion in the census does not influence inclusion in the P-sample or vice versa.

Two studies were directed at studying the adequacy of the homogeneity assumption. The first study (P13) compares the dual system estimates with demographic analysis to obtain an estimate of correlation bias at the national level. The second study (P17) is qualitative in nature, and compares the PES dual system estimates, the individual P- and E-samples, and demographic analysis to determine if inconsistencies exist that could indicate the presence of correlation bias due to failure of the homogeneity assumption.

The causality assumption is investigated by two qualitative studies (P14 a and b). The first of these studies pairs non-PES blocks with similar PES blocks and compares characteristics. There should be no difference between these blocks except for the random variation introduced by sampling. The second study uses a debriefing of field interviewers to assess the potential for correlation bias.

Project P13: Use of Alternative Dual System Estimators to Measure Correlation Bias

Alternative dual system estimators were developed using information from demographic analysis to try to address the problem of correlation bias due to failure of the homogeneity assumption—when people missed by the census are more likely to be missed by the PES than those included in the census and vice-versa. This was done by using demographic analysis sex ratios (the ratio of males to females) and the PES dual system estimates for females to create an alternative estimate for males. The DSE for females was multiplied by the sex ratio appropriate for each PES age group. By comparing these alternative estimates for males with the PES dual system estimates for males gives an estimate of correlation bias at the national level. The estimated correlation bias was then allocated to the individual PES male post-strata proportional to P-sample non-matches. This permitted estimates of correlation bias to be produced at the individual post-stratum level.

Project P17 Internal Consistency of Estimates

This study has two objectives: (1) to evaluate the reasonableness of the age sex distribution in the census and PES estimates and (2) to compare the PES and demographic analysis (DA) estimates of undercount to make some assessment of the accuracy of the PES estimates. For these purposes, sex ratios and information on undercount rates from the PES and DA were used. Sex ratio are used to evaluate if overall results on sex distribution are reasonable. Because demographic analysis estimates are available at the national level only, most comparison are limited to analyzing data for the U.S. by race black and non-black.

Project P14 Independence of the Census and P-Sample, Comparison of Blocks

The analysis for this project is directed at assessing the existence of correlation bias due to failure of the causality assumption:

The probability of an individual being included in the P-sample is not altered by inclusion in the census, and the probability of being included in the census is not altered by inclusion in the P-sample.

Several steps were implemented to study the existence of correlation bias. First, a sample of PES blocks paired with comparable non-PES blocks was drawn. The sample was selected by type of enumeration area (TEA) in order to do analyses isolating these groups. Each type of enumeration was analyzed as a separate data set since the timing of the PES and census operations were different across areas. Therefore, any PES effects on the census would be different for each TEA and should be tested using separate data sets.

The difference from PES blocks and non-PES blocks were the focus of the tests. For each block, relevant data were extracted from the final census files in January, 1991 and aggregated from person records to block level records. The preliminary variables were organized a priori into groups: block size, population coverage, housing unit status, mailback, field response, and edit & quality. The data were tested for relevance, completeness, and redundancy.

8. Small Area Estimation

Project P12: Evaluation of the Synthetic Assumption

Synthetic adjustment is used in the PES to "carry down" the estimated adjustment factors to the census counts in each post stratum. This synthetic adjustment assumes that the probability of being missed by the census is constant for each person within the post-stratum.

The coverage error may vary substantially within the PES strata although the post strata were drawn so as to be homogeneous with respect to expected coverage error. The goal of this study is to verify that the assumption underlying the synthetic adjustment is valid.

The analysis was based on studying the homogeneity of several different block level statistics. Three different types of analysis were conducted. First the distributions of census characteristics thought to be highly correlated with coverage error (e.g., mail return rate) were examined. Secondly, the distribution of the components of coverage error at the block level were studied. These components were erroneous enumeration rates and P-sample nonmatch rates. Finally, the production smoothing model was used to predict a block level adjustment factor for the same sample of blocks used for the first analysis.

The analysis concentrated on determining whether the block level statistics clustered unusually by state within the PES post-strata. Further analysis to examine clustering at other levels such as place and county remains to be carried out.

9. Late Late Census Data

Project P18: Evaluation of Late Late Census Data

Census data capture was completed after the completion of the last planned PES matching operation which was Late Census Data matching. A small amount of changes to census data (census additions, deletions and updated person data) resulted from the late census data capture activities. A portion of these changes were included into the PES results through the Late Late Census Data (LLCD) matching operation. The remainder of these late census data changes were not processed due to time constraints, and were not included in the PES results. The Evaluation of Late Late Census Data (Project 18) examines the effect that the late census data changes not included in the PES have on the PES estimates of undercount. The remaining late-late census data were processed to determine the effect that this would have had on the dual system estimates.

10. Total Error

Project 16: Total Error in PES Estimates for Evaluation Post Strata

The dual system estimator used in the estimation for the PES is known to be subject to various components of nonsampling error, in addition to sampling error. The PES evaluation program includes studies that provide direct measures of error due to nonsampling and sampling error components. These errors combine in the dual system estimator model to cause differences from population counts that would be attained under an error-free program. The difference between the PES estimate and the error-free count is referred to as the total error.

Project P16 evaluates both the components of error and the total error in the PES estimates for the 13 evaluation post strata. The components of error are response correlation bias (also called model bias), matching error, quality of reported Census Day address, fabrication in the P-sample, processing error in the E-sample, data collection error in the E-sample, error in balancing the estimates of the gross overcount and the gross undercount missing data (imputation error), sampling variance, and ratio estimator bias.

The evaluation of the total error assesses the overall accuracy of the PES estimates of population size and the census undercount rate. A synthesis of the components errors provides estimates of the bias and variance. This analysis then assesses the combined effect of the errors on the PES estimate of the undercount rate. The estimates of the mean and variance of the distributions of the component errors are based on the conclusions drawn from the various evaluation studies. The simulation method produced an estimate of the bias and variance of the estimated undercount rate.

The results of the total error model were also used in a loss function analysis to assess the accuracy of the distributions of population across states, places, and counties for the adjusted and unadjusted census. This analysis was carried out by forming target populations from the results of the total error work. The biases measured by the PES evaluations were incorporated into PES dual system estimates to produce corrected estimates of the population. These corrected estimates were designated as the target populations. The adjusted and unadjusted census population distributions were compared to the target population distributions using several loss functions. The comparisons were conducted at the state level and at the place and county level for the following size categories:

Places under 25,000 population.
Places of between 25,000 and 50,000 population.
Places of size over 50,000.
Counties under 200,000.
Counties larger than 200,000.

In addition, results were also produced for places and counties over 100,000 population.

Demographic Analysis

The Census Bureau's companion coverage measurement program to the PES was demographic analysis. The demographic coverage estimates could only be used to evaluate the completeness of coverage of the 1990 census at a national level and only for race (Black/Non-Black), sex, and age groups. Demographic analysis could not provide even reasonably reliable coverage estimates for the Hispanic, Asian/Pacific Islander, or American Indian/Native Alaskan populations because these characteristics have not always been recorded on birth and death certificates; nor can the demographic method provide direct estimates of the resident population at the State or substate level. However, the PES measured under or overcounts of these groups. The demographic coverage estimates were compared to the post-enumeration survey coverage estimates to assess the overall consistency of the two sets of estimates at the national level.

Demographic analysis uses historical data on births, deaths, and legal immigration; estimates of emigration and undocumented immigration; and Medicare data to develop an independent estimate of the resident population on census day. The estimate is compared with the census count to yield a measure of net census coverage and net undercount. The particular procedure that is used to estimate coverage nationally in 1990 for the various demographic subgroups depends primarily on the nature and availability of the required demographic data. Birth and death records are available for the entire United States from 1933 on for developing estimates of population at ages under 57 in 1990. In estimating births for each year, the Bureau added to the number of registered births an estimate of underregistration. Underregistration was estimated based on tests conducted in 1940, 1950, and 1964–1968. If the estimates of underregistration are off, they could have a significant effect on undercount estimates because birth data are by far the largest component in estimating the population through demographic analysis. In fact, in producing the demographic estimates of population for 1990 the Bureau revised the estimates for certain Black birth cohorts to account for biases that recent research identified in the birth registration test result of 1940.

National birth and death records are not available before 1933, so the Bureau had to find other ways to estimate the population size of these cohorts in 1990 (ages 55 and over were estimated). For the population 65 and over, administrative data on aggregate Medicare enrollments for 1990 (adjusted for underenrollment) are used to estimate population and net coverage. For the Non-black population aged 55 to 64 in 1990, the estimates of population are based primarily on national birth estimates for 1925–1934 developed by Whelpton. For the Black population aged 55 to 64 in 1990, the estimates of population are based on revisions of estimates for the cohort in 1960 developed by Coale and Rives.

In addition to subtracting deaths, the estimates of births described above are augmented to account for change due to immigration, emigration, and net international movement abroad of citizens (including the Armed Forces and Puerto Rican migrants). The various components of net migration vary significantly in their completeness and quality. The United States does not keep emigration records. Therefore, an estimate had to be made of those who have left the country. While the United States does have good records of legal

immigration, there is no accurate estimate of illegal immigration—the most elusive demographic component of population change. The Census Bureau has developed a preliminary estimate for undocumented residents in 1990 based on analysis of survey data and administrative records of the Immigration and Naturalization Service (INS). The INS now collects different information than it did prior to 1980. Recent immigration reform further complicated the effort to estimate legal immigration and undocumented residents. Although the legislative reform allowed many undocumented aliens to receive amnesty, some of these persons may not actually reside in the United States.

It should be noted that before the demographic estimates of population for race groups are compared to the census to calculate the net undercount, the race categories of the census counts must be "modified" so that they are consistent with the race categories of the historical demographic estimates. Specifically, 9.8 million persons in the 1990 census (mostly of Hispanic origin) reported their race in the "Other race-not specified" category, a category not included in the demographic estimates. This modification added 497,000 persons to the census count for Blacks. Also, the age categories of the 1990 census counts have been "modified" so they are consistent with the April 1, 1990 time reference of the demographic estimates.

It is important to emphasize that results of demographic analysis are not exact but are estimates. To a large extent, they were based on assumptions and best professional judgment. As in the PES, the Bureau tried to estimate potential error in the data produced by demographic analysis. To estimate that overall error, the Bureau conducted 11 detailed demographic analysis evaluation studies to find out as much as possible about each possible source of error—the specific projects are identified in Table 1. Based on these studies, the Bureau developed a range of error around the demographic analysis estimates. Since these evaluation projects and the demographic error model represent an evaluation program new for the 1990 census, the assessments of potential error are subject to change and improvement over time just as the basic demographic estimates of coverage have been.

Table 1.— The Eleven Demographic Analysis Evaluation Projects

D1..... Error in Birth Underregistration Completeness Estimates.

Table 1.— The Eleven Demographic Analysis Evaluation Projects—Continued

D2..... Uncertainty in Estimates of Undocumented Aliens.
D3..... Uncertainty in Estimated White Births, 1915–1935.
D4..... Uncertainty in Estimated Black Births, 1915–1935.
D5..... Robustness of Estimated Number of Emigrants.
D6..... Robustness of Estimates of the Population 65 and Older.
D7..... Uncertainty Measures for Other Components.
D8..... Uncertainty of Models to Translate 1990 Census Concepts into Historical Racial Classifications.
D9..... Inconsistencies in Race Classifications of the Demographic Estimates and the Census.
D10..... Differences Between Preliminary and Final Demographic Estimates.
D11..... Total Error in the Demographic Estimates.

Attachment 1

PES SAMPLE SIZE BY STATE (P-SAMPLE)

State names	Blocks	Clusters	Housing units
Alabama	280	168	4,706
Alaska	27	16	946
Arizona	569	115	5,046
Arkansas	161	77	2,230
California	652	390	13,013
Colorado	401	101	3,290
Connecticut	74	55	1,816
Delaware	19	12	460
District of Columbia	22	18	657
Florida	298	198	5,973
Georgia	189	112	3,320
Hawaii	49	19	599
Idaho	226	51	1,697
Illinois	300	221	7,553
Indiana	149	92	2,540
Iowa	179	86	2,491
Kansas	264	74	2,188
Kentucky	177	107	3,116
Louisiana	165	105	3,481
Maine	216	67	2,292
Maryland	72	56	2,162
Massachusetts	162	107	3,185
Michigan	232	152	4,959
Minnesota	256	99	3,186
Mississippi	179	103	2,696
Missouri	215	116	3,369
Montana	409	46	1,755
Nebraska	140	44	1,257
Nevada	66	27	1,195
New Hampshire	118	49	1,987
New Jersey	117	91	2,752
New Mexico	553	68	2,533
New York	520	371	12,210
North Carolina	209	126	3,754
North Dakota	205	19	679
Ohio	216	146	4,491
Oklahoma	271	93	2,737
Oregon	310	83	2,575
Pennsylvania	499	303	9,517
Rhode Island	32	24	832
South Carolina	107	58	1,900
South Dakota	230	18	686
Tennessee	243	173	4,858
Texas	845	436	12,807
Utah	212	40	1,351
Vermont	115	28	1,423
Virginia	144	87	2,609
Washington	352	111	3,939
West Virginia	49	31	911
Wisconsin	141	76	2,264
Wyoming	488	26	801
National Total	12,124	5,290	168,794

Attachment 2

1990 Post-Enumeration Survey Post Strata

The 1990 Post-Enumeration Survey (PES) will provide direct estimates for 1392 post strata. The post strata are designed to divide the PES sample blocks into groups which have similar characteristics. This helps the Census Bureau to estimate the coverage of the 1990 decennial census more accurately.

The post strata are defined by census division, area (city, non-city, rural, etc.), race, Hispanic origin, tenure group, sex, and age. Tenure refers to whether housing units are owned or rented. Each post strata is given an eight digit code. The attached document shows 116 post strata and the corresponding first six digits of the post stratum code for each. The last two digits are not delineated on the attachment. They define sex and age group. There are six age group classifications. What follows is an explanation of the post strata coding system:

The first digit of each given eight digit code defines the census division. The nine census divisions and the states in each census division are:

1—New England—Connecticut, Massachusetts, Maine, New Hampshire, Rhode Island, and Vermont
2—Middle Atlantic—New Jersey, New York, and Pennsylvania
3—South Atlantic—Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, and West Virginia
4—East South Central—Alabama, Kentucky, Mississippi, Tennessee
5—West South Central—Arkansas, Louisiana, Oklahoma, and Texas
6—East North Central—Illinois, Indiana, Michigan, Ohio, and Wisconsin
7—West North Central—Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota
8—Mountain—Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming

9— Pacific—Alaska, California, Hawaii, Oregon, and Washington

Within each census division, the geographic areas are divided by type of area. There are nine possible type of area codes:

0—Central cities in explicitly named PMSAs (see description below)
1—Central cities in large metropolitan areas (Type I MAs)
2—Central cities in small metropolitan areas (Type II MAs)
3—Central cities in a metropolitan area regardless of size
4—Non-central city areas in the New York PMSA
5—Non-central city areas in large metropolitan areas (Type I MAs)
6—Non-central city areas in small metropolitan areas (Type II MAs)
7—Non-central city areas in metropolitan areas
8—Non-metropolitan areas incorporated places with 10,000 + population
9—Balance of non-metropolitan areas

A PMSA is a Primary Metropolitan Statistical Area. There are four explicitly named PMSAs in the 1990 PES post strata. These PMSAs and the census division in which they are located are:

- The New York City PMSA in the Middle Atlantic division,
- The Houston PMSA plus the Dallas PMSA, plus the Fort Worth PMSA in the West South Central division,
- The Chicago PMSA plus the Detroit PMSA in the East North Central division,
- The Los Angeles-Long Beach PMSA in the Pacific division.

A large metropolitan area (type I MA) is an area whose largest central city has a population of at least 250,000 using the 1990 census person count.

A small metropolitan area (type II MA) is an area which does not have any central cities with a population of 250,000 or more.

The balance of non-metropolitan areas consist of areas which are not included in area type number 8. This would consist primarily of rural areas.

Any post strata can include up to three area types. The area types included in a stratum are delineated in the second to fourth digits of the post strata code. For instance, post strata code 578910 includes area types 7, 8, and 9. But most post strata contain only one area type. If a post stratum has only one area type, the second digit of the post stratum code indicates the area type, and the third and fourth digits are zero. In general, each of the second through fourth digits is filled with a zero from the right if a given geographic area of post stratum contains less than three area types.

The race/hispanic origin is determined by the fifth digit of the post stratum code. The tenure group is determined by the sixth digit of the post stratum code. These three attributes are combined in the coding system. The possible race/hispanic origin groups are: Black, Non-Black Hispanic, Asian-Pacific Islander, American Indian, and Other. A post stratum can consist of more than one race/hispanic origin group. This is reflected in the definitions below. The tenure designation defines whether the persons in the geographic area are owners or renters. Some geographic areas were not divided by tenure. The possible codes for the fifth and sixth digits are:

10—Black (Renter & Owner)
11—Black Renter
12—Black Owner
20—Non-Black Hispanic (Renter & Owner)
21—Non-Black Hispanic Renter
22—Non-Black Hispanic Owner
30—All Other (Renter & Owner)
31—All Other Renter
32—All Other Owner
40—Asian-Pacific Islander (Renter & Owner)
41—Asian-Pacific Islander Renter
42—Asian-Pacific Islander Owner
50—Black and Non-Black Hispanic (Renter & Owner) & Non-Black Non-Asian-Pacific Islander Hispanic
60—American Indian

The seventh digit of the post stratum code defines the sex.

1—Male
2—Female

Within sex there are six age groups, the eighth digit. The age groups are:

1—0–9
2—10–19
3—20–29
4—30–44
5—45–64
6—65+

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]

Stratum code	Factor
09006011	1.166
09006012	1.182
09006013	1.158
09006014	1.197
09006015	1.117
09006016	1.143
09006021	1.130
09006022	1.189
09006023	1.166
09006024	1.071
09006025	1.068
09006026	1.097
13003011	1.001
13003012	0.987
13003013	1.034
13003014	0.984
13003015	0.991
13003016	0.964
13003021	0.989
13003022	0.979
13003023	1.007
13003024	0.976
13003025	0.981

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
13003026	0.957
13705011	1.068
13705012	1.027
13705013	1.079
13705014	1.068
13705015	1.040
13705016	1.012
13705021	1.047
13705022	1.003
13705023	1.050
13705024	1.041
13705025	1.012
13705026	1.015
17003011	1.020
17003012	0.989
17003013	1.030
17003014	0.990
17003015	1.014
17003016	0.987
17003021	1.016
17003022	0.974
17003023	1.021
17003024	1.007
17003025	0.994
17003026	0.975
18003011	1.025
18003012	0.980
18003013	1.030
18003014	1.028
18003015	1.011
18003016	0.984
18003021	1.007
18003022	0.974
18003023	1.003
18003024	1.008
18003025	1.002
18003026	0.995
19003011	1.022
19003012	1.008
19003013	1.073
19003014	1.028
19003015	1.024
19003016	1.013
19003021	1.017
19003022	1.003
19003023	1.018
19003024	1.013
19003025	0.996
19003026	1.006
20001111	1.111
20001112	1.076
20001113	1.122
20001114	1.102
20001115	1.043
20001116	1.077
20001121	1.112
20001122	1.031
20001123	1.090
20001124	1.114
20001125	1.038
20001126	1.050
20001211	1.022
20001212	0.994
20001213	1.010
20001214	0.990
20001215	0.991
20001216	0.980
20001221	1.055
20001222	0.997
20001223	1.019
20001224	0.989
20001225	0.982
20001226	0.981
20002011	1.050
20002012	0.990
20002013	1.053
20002014	1.018
20002015	1.024

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
20002016	1.002
20002021	0.995
20002022	1.002
20002023	1.033
20002024	1.015
20002025	1.005
20002026	0.994
20003111	0.993
20003112	0.997
20003113	1.113
20003114	1.041
20003115	1.016
20003116	0.964
20003121	1.001
20003122	0.954
20003123	1.054
20003124	1.011
20003125	0.987
20003126	0.935
20003211	0.988
20003212	0.993
20003213	1.013
20003214	1.001
20003215	1.017
20003216	0.954
20003221	1.030
20003222	0.980
20003223	1.017
20003224	1.012
20003225	0.972
20003226	1.002
20004011	1.130
20004012	1.124
20004013	1.156
20004014	1.107
20004015	1.104
20004016	1.095
20004021	1.128
20004022	1.069
20004023	1.130
20004024	1.133
20004025	1.101
20004026	1.081
21001111	1.092
21001112	1.037
21001113	1.126
21001114	1.107
21001115	1.063
21001116	1.033
21001121	1.090
21001122	1.076
21001123	1.127
21001124	1.083
21001125	1.055
21001126	1.035
21001211	1.022
21001212	1.040
21001213	1.029
21001214	0.989
21001215	0.992
21001216	0.988
21001221	1.037
21001222	0.984
21001223	1.010
21001224	0.969
21001225	0.986
21001226	0.989
21003111	1.002
21003112	0.970
21003113	1.034
21003114	1.003
21003115	0.969
21003116	0.984
21003121	0.991
21003122	0.997
21003123	1.001
21003124	1.008
21003125	0.985
21003126	0.959
21003211	1.024
21003212	0.956
21003213	1.013
21003214	1.020
21003215	0.998
21003216	0.982
21003221	1.005
21003222	0.995
21003223	1.049
21003224	0.987
21003225	0.991
21003226	0.979
22001011	1.127
22001012	1.031
22001013	1.129
22001014	1.142
22001015	1.103
22001016	0.057
22001021	1.157
22001022	1.080
22001023	1.140
22001024	1.071
22001025	1.074
22001026	1.058
22003011	0.991
22003012	0.989
22003013	1.037
22003014	1.022
22003015	1.017
22003016	0.975
22003021	1.008
22003022	0.968
22003023	1.002
22003024	0.993
22003025	1.009
22003026	0.974
23002011	1.010
23002012	1.021
23002013	1.071
23002014	1.022
23002015	1.008
23002016	0.972
23002021	1.024
23002022	0.976
23002023	1.008
23002024	1.055
23002025	1.010
23002026	0.995
24003011	1.053
24003012	0.991
24003013	1.020
24003014	1.012
24003015	0.996
24003016	0.981
24003021	1.017
24003022	1.006
24003023	1.057
24003024	0.991
24003025	0.979
24003026	0.978
24505011	1.071
24505012	1.057
24505013	1.115
24505014	1.095
24505015	1.060
24505016	1.060
24505021	1.108
24505022	1.032
24505023	1.063
24505024	1.085
24505025	1.057
24505026	1.014
25003011	1.009
25003012	0.983
25003013	1.037
25003014	1.031
25003015	0.981
25003016	0.971
25003021	1.029
25003022	1.018
25003023	1.033
25003024	1.002
25003025	0.983
25003026	0.973
26003011	1.018
26003012	0.990
26003013	1.040
26003014	0.994
26003015	0.991
26003016	0.984
26003021	1.003
26003022	0.978
26003023	0.999
26003024	1.011
26003025	0.994
26003026	0.984
28003011	1.015
28003012	0.987
28003013	1.017
28003014	1.030
28003015	1.006
28003016	0.991
28003021	1.061
28003022	0.975
28003023	1.016
28003024	0.998
28003025	0.992
28003026	0.984
29003011	1.014
29003012	0.991
29003013	1.042
29003014	1.019
29003015	0.993
29003016	1.001
29003021	1.009
29003022	0.999
29003023	1.041
29003024	1.017
29003025	0.982
29003026	0.986
29995011	1.071
29995012	1.048
29995013	1.067
29995014	1.074
29995015	1.037
29995016	1.033
29995021	1.055
29995022	1.045
29995023	1.054
29995024	1.068
29995025	1.054
29995026	1.039
31001111	1.133
31001112	1.102
31001113	1.106
31001114	1.131
31001115	1.076
31001116	1.086
31001121	1.155
31001122	1.096
31001123	1.105
31001124	1.069
31001125	1.067
31001126	1.037
31001211	1.066
31001212	1.017
31001213	1.030
31001214	1.024
31001215	0.990
31001216	0.991
31001221	1.037
31001222	1.008
31001223	1.027
31001224	0.992
31001225	0.994

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
31001226	0.977
31003111	1.085
31003112	1.038
31003113	1.073
31003114	1.065
31003115	1.047
31003116	0.993
31003121	1.054
31003122	1.055
31003123	1.099
31003124	1.013
31003125	1.011
31003126	0.983
31003211	1.039
31003212	1.035
31003213	1.048
31003214	1.035
31003215	0.983
31003216	0.985
31003221	1.035
31003222	1.031
31003223	1.073
31003224	1.021
31003225	0.979
31003226	1.008
32001011	1.052
32001012	1.035
32001013	1.072
32001014	1.037
32001015	1.015
32001016	1.006
32001021	1.084
32001022	1.028
32001023	1.083
32001024	1.047
32001025	1.003
32001026	0.981
32003011	1.065
32003012	1.068
32003013	1.080
32003014	1.046
32003015	1.027
32003016	0.986
32003021	1.048
32003022	1.032
32003023	1.039
32003024	1.007
32003025	0.987
32003026	0.998
33002011	1.106
33002012	1.064
33002013	1.101
33002014	1.088
33002015	1.005
33002016	0.985
33002021	1.101
33002022	1.056
33002023	1.091
33002024	1.065
33002025	0.984
33002026	0.984
35001011	1.042
35001012	1.012
35001013	1.034
35001014	1.007
35001015	0.996
35001016	0.990
35001021	1.040
35001022	1.012
35001023	1.045
35001024	1.017
35001025	1.007
35001026	0.986
35003011	1.030
35003012	0.997
35003013	1.032
35003014	1.008
35003015	0.982

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
35003016	0.985
35003021	1.036
35003022	1.015
35003023	1.035
35003024	0.995
35003025	0.975
35003026	0.983
36001011	1.074
36001012	1.033
36001013	1.034
36001014	1.044
36001015	1.018
36001016	1.003
36001021	1.035
36001022	1.043
36001023	1.051
36001024	1.042
36001025	1.010
36001026	1.001
36003011	1.052
36003012	1.007
36003013	1.039
36003014	1.042
36003015	0.991
36003016	0.992
36003021	1.069
36003022	1.062
36003023	1.038
36003024	1.043
36003025	1.028
36003026	0.994
37892011	1.030
37892012	1.083
37892013	1.133
37892014	1.074
37892015	1.007
37892016	1.017
37892021	1.090
37892022	1.021
37892023	1.068
37892024	1.059
37892025	0.994
37892026	0.971
38001011	1.025
38001012	1.001
38001013	1.023
38001014	1.033
38001015	1.023
38001016	0.984
38001021	1.057
38001022	1.015
38001023	1.048
38001024	1.021
38001025	0.953
38001026	0.963
38003011	1.058
38003012	1.015
38003013	1.066
38003014	1.020
38003015	1.000
38003016	0.981
38003021	1.046
38003022	1.010
38003023	1.026
38003024	1.007
38003025	0.995
38003026	0.979
39001011	1.057
39001012	1.039
39001013	1.021
39001014	1.039
39001015	1.023
39001016	0.981
39001021	1.071
39001022	1.045
39001023	1.045
39001024	0.999
39001025	0.994

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
39001026	0.979
39003011	1.047
39003012	1.014
39003013	1.063
39003014	1.035
39003015	1.002
39003016	0.994
39003021	1.058
39003022	1.045
39003023	1.060
39003024	1.022
39003025	1.022
39003026	0.997
41003111	1.064
41003112	1.074
41003113	1.056
41003114	1.078
41003115	1.015
41003116	0.989
41003121	1.075
41003122	1.050
41003123	1.042
41003124	1.062
41003125	1.025
41003126	0.982
41003211	1.045
41003212	1.032
41003213	1.042
41003214	1.043
41003215	1.011
41003216	0.994
41003221	1.065
41003222	1.028
41003223	1.064
41003224	1.038
41003225	1.006
41003226	1.001
42003011	1.075
42003012	1.018
42003013	1.072
42003014	1.042
42003015	1.002
42003016	0.996
42003021	1.036
42003022	1.055
42003023	1.058
42003024	1.020
42003025	0.987
42003026	0.975
43005011	1.093
43005012	1.075
43005013	1.090
43005014	1.085
43005015	1.055
43005016	1.009
43005021	1.116
43005022	1.041
43005023	1.083
43005024	1.043
43005025	1.003
43005026	0.987
47003011	1.041
47003012	1.023
47003013	1.043
47003014	1.042
47003015	1.002
47003016	0.987
47003021	1.051
47003022	1.024
47003023	1.050
47003024	1.015
47003025	1.007
47003026	0.990
47895011	1.062
47895012	1.008
47895013	1.020
47895014	1.042
47895015	1.004

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
47895016	0.999
47895021	1.050
47895022	1.011
47895023	1.038
47895024	1.027
47895025	1.004
47895026	0.975
48003011	1.039
48003012	1.029
48003013	1.053
48003014	1.020
48003015	1.003
48003016	0.996
48003021	1.048
48003022	1.017
48003023	1.032
48003024	1.014
48003025	0.988
48003026	1.004
49003011	1.032
49003012	1.021
49003013	1.066
49003014	1.012
49003015	0.990
49003016	0.998
49003021	1.032
49003022	1.016
49003023	1.060
49003024	1.010
49003025	0.987
49003026	1.010
50001011	1.098
50001012	1.081
50001013	1.096
50001014	1.072
50001015	1.042
50001016	1.020
50001021	1.104
50001022	1.057
50001023	1.117
50001024	1.058
50001025	1.022
50001026	0.995
50002011	1.088
50002012	1.044
50002013	1.143
50002014	1.063
50002015	1.014
50002016	0.963
50002021	1.128
50002022	1.079
50002023	1.105
50002024	1.043
50002025	0.992
50002026	0.958
50003111	1.058
50003112	1.050
50003113	1.073
50003114	1.060
50003115	1.035
50003116	1.008
50003121	1.080
50003122	1.043
50003123	1.053
50003124	1.019
50003125	1.028
50003126	0.990
50003211	1.033
50003212	1.004
50003213	1.054
50003214	1.020
50003215	1.017
50003216	0.977
50003221	1.033
50003222	1.019
50003223	1.027
50003224	1.025
50003225	0.997

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
50003226	0.999
51003111	1.041
51003112	1.038
51003113	1.027
51003114	1.032
51003115	1.014
51003116	0.982
51003121	1.059
51003122	1.039
51003123	1.069
51003124	1.028
51003125	0.999
51003126	0.980
51003211	1.032
51003212	1.014
51003213	1.039
51003214	1.012
51003215	0.994
51003216	0.984
51003221	1.027
51003222	1.011
51003223	1.041
51003224	1.005
51003225	0.994
51003226	0.985
52003011	1.053
52003012	1.034
52003013	1.056
52003014	1.038
52003015	1.003
52003016	0.995
52003021	1.045
52003022	1.017
52003023	1.053
52003024	1.027
52003025	0.994
52003026	0.988
53001011	1.079
53001012	1.034
53001013	1.099
53001014	1.057
53001015	1.032
53001016	1.001
53001021	1.089
53001022	1.047
53001023	1.074
53001024	1.045
53001025	0.989
53001026	0.997
53002011	1.065
53002012	1.037
53002013	1.051
53002014	1.033
53002015	0.974
53002016	0.970
53002021	1.095
53002022	1.023
53002023	1.094
53002024	1.048
53002025	0.971
53002026	0.979
57003011	1.044
57003012	1.043
57003013	1.050
57003014	1.024
57003015	0.999
57003016	0.989
57003021	1.033
57003022	1.032
57003023	1.048
57003024	1.030
57003025	0.993
57003026	0.971
57891011	1.069
57891012	1.016
57891013	1.041
57891014	1.033
57891015	1.004

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
57891016	0.988
57891021	1.032
57891022	1.003
57891023	1.060
57891024	1.013
57891025	0.992
57891026	0.985
57892011	1.056
57892012	1.058
57892013	1.077
57892014	1.065
57892015	1.047
57892016	1.002
57892021	1.081
57892022	1.047
57892023	1.059
57892024	1.043
57892025	1.011
57892026	1.015
58003011	1.027
58003012	0.994
58003013	1.048
58003014	1.030
58003015	0.999
58003016	0.978
58003021	1.027
58003022	1.021
58003023	1.062
58003024	1.013
58003025	0.993
58003026	0.969
59003011	1.052
59003012	1.032
59003013	1.029
59003014	1.019
59003015	1.014
59003016	0.990
59003021	1.043
59003022	1.016
59003023	1.043
59003024	1.034
59003025	1.006
59003026	0.983
60001111	1.131
60001112	1.031
60001113	1.067
60001114	1.081
60001115	1.042
60001116	0.984
60001121	1.112
60001122	1.061
60001123	1.106
60001124	1.020
60001125	1.004
60001126	0.954
60001211	1.047
60001212	1.001
60001213	1.042
60001214	1.039
60001215	1.015
60001216	1.029
60001221	1.050
60001222	1.040
60001223	1.033
60001224	1.007
60001225	0.994
60001226	0.975
60003111	1.016
60003112	1.021
60003113	1.087
60003114	1.093
60003115	1.108
60003116	1.028
60003121	1.111
60003122	1.052
60003123	1.123
60003124	0.989
60003125	1.017

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
60003126	0.924
60003211	1.021
60003212	1.010
60003213	1.001
60003214	1.027
60003215	1.016
60003216	1.005
60003221	1.047
60003222	1.005
60003223	1.031
60003224	1.000
60003225	1.000
60003226	1.000
60102011	0.958
60102012	1.005
60102013	1.000
60102014	1.026
60102015	0.991
60102016	0.998
60102021	0.993
60102022	0.989
60102023	0.966
60102024	0.957
60102025	0.936
60102026	0.963
61001111	1.042
61001112	1.003
61001113	1.100
61001114	1.052
61001115	1.034
61001116	0.999
61001121	1.086
61001122	1.072
61001123	1.058
61001124	1.047
61001125	0.994
61001126	0.955
61001211	1.091
61001212	0.983
61001213	1.025
61001214	1.026
61001215	1.006
61001216	1.007
61001221	1.045
61001222	1.014
61001223	1.012
61001224	1.000
61001225	0.949
61001226	0.977
61003111	1.119
61003112	0.954
61003113	0.992
61003114	1.070
61003115	1.033
61003116	0.970
61003121	1.030
61003122	0.980
61003123	1.010
61003124	0.999
61003125	0.940
61003126	0.972
61003211	0.957
61003212	0.971
61003213	1.036
61003214	1.021
61003215	0.973
61003216	0.994
61003221	0.986
61003222	0.989
61003223	1.011
61003224	1.003
61003225	1.016
61003226	1.001
62003011	1.033
62003012	0.978
62003013	1.064
62003014	1.019
62003015	1.016
62003016	1.020
62003021	1.022
62003022	1.012
62003023	1.032
62003024	1.032
62003025	1.002
62003026	1.008
62705011	1.088
62705012	1.054
62705013	1.090
62705014	1.062
62705015	1.021
62705016	1.006
62705021	1.095
62705022	1.066
62705023	1.074
62705024	1.030
62705025	1.024
62705026	1.002
65003011	1.017
65003012	0.999
65003013	1.030
65003014	1.014
65003015	1.002
65003016	1.003
65003021	1.011
65003022	1.012
65003023	1.008
65003024	0.995
65003025	0.992
65003026	0.997
66003011	1.017
66003012	0.988
66003013	1.023
66003014	1.034
66003015	0.999
66003016	0.972
66003021	1.013
66003022	1.008
66003023	1.000
66003024	1.018
66003025	0.978
66003026	1.002
68003011	1.005
68003012	0.977
68003013	1.028
68003014	1.008
68003015	1.008
68003016	0.997
68003021	1.003
68003022	1.005
68003023	1.019
68003024	0.997
68003025	0.988
68003026	0.987
69003011	0.991
69003012	0.981
69003013	1.019
69003014	0.987
69003015	0.986
69003016	0.997
69003021	0.984
69003022	0.981
69003023	1.014
69003024	0.982
69003025	0.982
69003026	0.995
71003111	1.064
71003112	0.995
71003113	1.107
71003114	1.054
71003115	1.041
71003116	0.997
71003121	1.007
71003122	1.015
71003123	1.012
71003124	0.981
71003125	0.996
71003126	0.940
71003211	0.987
71003212	0.980
71003213	1.021
71003214	0.992
71003215	0.996
71003216	0.981
71003221	1.005
71003222	1.026
71003223	1.004
71003224	0.994
71003225	1.000
71003226	0.988
71005011	1.110
71005012	1.030
71005013	1.080
71005014	1.049
71005015	1.037
71005016	1.011
71005021	1.095
71005022	1.072
71005023	1.077
71005024	1.035
71005025	1.022
71005026	1.009
72003011	1.003
72003012	1.006
72003013	1.060
72003014	1.011
72003015	1.003
72003016	1.010
72003021	1.038
72003022	0.990
72003023	1.067
72003024	1.007
72003025	1.001
72003026	1.014
72505011	1.116
72505012	1.045
72505013	1.101
72505014	1.091
72505015	1.038
72505016	1.021
72505021	1.114
72505022	1.068
72505023	1.084
72505024	1.091
72505025	1.027
72505026	1.023
75003011	1.011
75003012	1.013
75003013	1.028
75003014	1.006
75003015	1.001
75003016	0.999
75003021	1.025
75003022	0.990
75003023	1.027
75003024	0.993
75003025	1.003
75003026	0.996
76003011	1.030
76003012	0.988
76003013	1.056
76003014	1.022
76003015	1.002
76003016	1.021
76003021	1.023
76003022	1.028
76003023	1.020
76003024	1.007
76003025	1.000
76003026	1.021
78003011	1.003
78003012	0.985
78003013	1.023
78003014	1.025
78003015	1.008

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
78003016	1.001
78003021	1.041
78003022	0.987
78003023	1.016
78003024	1.016
78003025	0.997
78003026	0.981
79003011	1.013
79003012	0.995
79003013	1.041
79003014	0.999
79003015	0.993
79003016	1.001
79003021	1.010
79003022	1.000
79003023	1.015
79003024	0.994
79003025	1.001
79003026	0.997
79995011	1.077
79995012	1.007
79995013	1.082
79995014	1.065
79995015	1.033
79995016	1.021
79995021	1.069
79995022	1.026
79995023	1.725
79995024	1.048
79995025	1.027
79995026	0.989
81003111	1.034
81003112	1.035
81003113	1.123
81003114	1.086
81003115	1.041
81003116	0.990
81003121	1.041
81003122	1.061
81003123	1.060
81003124	1.019
81003125	0.977
81003126	1.013
81003211	1.031
81003212	1.021
81003213	1.035
81003214	1.020
81003215	1.002
81003216	0.986
81003221	1.031
81003222	1.020
81003223	1.045
81003224	1.003
81003225	0.990
81003226	0.980
82003011	1.017
82003012	1.017
82003013	1.071
82003014	1.014
82003015	0.978
82003016	0.975
82003021	1.030
82003022	1.021
82003023	1.064
82003024	0.998
82003025	0.987
82003026	0.982
83005011	1.066
83005012	1.023
83005013	1.107
83005014	1.063
83005015	1.027
83005016	1.005
83005021	1.058
83005022	1.055
83005023	1.077
83005024	1.025
83005025	0.973

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
83005026	1.004
87003011	1.022
87003012	1.008
87003013	1.076
87003014	1.023
87003015	0.988
87003016	0.986
87003021	1.026
87003022	0.984
87003023	1.022
87003024	1.000
87003025	0.971
87003026	0.985
88003011	1.021
88003012	1.032
88003013	1.059
88003014	1.026
88003015	0.990
88003016	0.993
88003021	1.036
88003022	1.028
88003023	1.036
88003024	0.996
88003025	0.971
88003026	0.995
89003011	1.050
89003012	1.027
89003013	1.077
89003014	1.036
89003015	1.031
89003016	1.003
89003021	1.046
89003022	1.049
89003023	1.041
89003024	1.024
89003025	1.017
89003026	1.014
89995011	1.110
89995012	1.076
89995013	1.123
89995014	1.070
89995015	1.039
89995016	1.062
89995021	1.106
89995022	1.084
89995023	1.105
89995024	1.057
89995025	1.076
89995026	1.049
90003111	1.043
90003112	1.076
90003113	1.098
90003114	1.094
90003115	1.004
90003116	0.963
90003121	1.047
90003122	1.056
90003123	1.089
90003124	1.014
90003125	1.015
90003126	0.977
90003211	1.017
90003212	1.043
90003213	1.022
90003214	1.011
90003215	1.000
90003216	1.012
90003221	1.037
90003222	1.041
90003223	1.019
90003224	1.031
90003225	1.000
90003226	0.995
90301111	1.142
90301112	1.075
90301113	1.115
90301114	1.124
90301115	1.127

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
90301116	0.971
90301121	1.105
90301122	1.020
90301123	1.089
90301124	1.008
90301125	0.971
90301126	0.960
90301211	1.160
90301212	1.056
90301213	1.128
90301214	1.116
90301215	1.094
90301216	1.013
90301221	1.132
90301222	1.070
90301223	1.108
90301224	1.074
90301225	1.042
90301226	1.054
90302111	1.093
90302112	1.055
90302113	1.159
90302114	1.095
90302115	1.088
90302116	0.992
90302121	1.055
90302122	1.080
90302123	1.106
90302124	1.035
90302125	1.030
90302126	1.007
90302211	1.052
90302212	1.000
90302213	1.042
90302214	1.004
90302215	0.990
90302216	0.990
90302221	1.029
90302222	1.016
90302223	1.094
90302224	1.025
90302225	0.971
90302226	0.995
90304111	1.047
90304112	1.053
90304113	1.147
90304114	1.090
90304115	1.074
90304116	1.046
90304121	1.069
90304122	1.060
90304123	1.068
90304124	1.074
90304125	0.981
90304126	1.057
90304211	1.076
90304212	1.052
90304213	1.071
90304214	1.065
90304215	1.028
90304216	1.029
90304221	1.070
90304222	1.072
90304223	1.079
90304224	1.026
90304225	1.033
90304226	1.029
91003111	1.035
91003112	1.045
91003113	1.112
91003114	1.073
91003115	1.020
91003116	0.986
91003121	1.000
91003122	1.033
91003123	1.045
91003124	1.020
91003125	0.947

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
91003126	0.959
91003211	1.025
91003212	0.978
91003213	1.017
91003214	1.020
91003215	1.004
91003216	0.999
91003221	1.038
91003222	1.030
91003223	1.092
91003224	1.023
91003225	1.001
91003226	0.988
92003011	1.041
92003012	1.005
92003013	1.070
92003014	1.016
92003015	0.992
92003016	0.983
92003021	0.987
92003022	1.011
92003023	1.028
92003024	1.010
92003025	0.985
92003026	0.973
95003011	1.028
95003012	0.995
95003013	1.050
95003014	0.971
95003015	1.001
95003016	0.979
95003021	1.051
95003022	1.002
95003023	1.029
95003024	0.995
95003025	0.987
95003026	0.967
96003011	1.026
96003012	1.039
96003013	1.014
96003014	1.018
96003015	1.070
96003016	1.002

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
96003021	1.043
96003022	1.108
96003023	1.065
96003024	1.030
96003025	1.010
96003026	0.976
97001011	1.251
97001012	1.235
97001013	1.250
97001014	1.278
97001015	1.180
97001016	1.117
97001021	1.199
97001022	1.158
97001023	1.182
97001024	1.136
97001025	1.111
97001026	1.112
97002011	1.092
97002012	1.084
97002013	1.088
97002014	1.085
97002015	1.048
97002016	1.014
97002021	1.088
97002022	1.071
97002023	1.079
97002024	1.052
97002025	1.061
97002026	1.018
97004011	1.026
97004012	0.992
97004013	1.048
97004014	1.009
97004015	0.993
97004016	0.963
97004021	1.063
97004022	0.990
97004023	1.053
97004024	1.011
97004025	0.932
97004026	0.974
98003011	1.044

ATTACHMENT 3.—ADJUSTMENT FACTORS BY POST STRATUM [1]—Continued

Stratum code	Factor
98003012	1.027
98003013	1.053
98003014	1.009
98003015	0.996
98003016	0.967
98003021	1.052
98003022	1.024
98003023	1.045
98003024	1.010
98003025	0.994
98003026	1.009
98904011	0.995
98904012	1.008
98904013	1.033
98904014	1.029
98904015	0.985
98904016	0.973
98904021	0.996
98904022	1.013
98904023	1.025
98904024	1.008
98904025	0.985
98904026	0.942
99003011	1.028
99003012	1.036
99003013	1.043
99003014	1.024
99003015	1.005
99003016	0.994
99003021	1.029
99003022	1.020
99003023	1.048
99003024	1.019
99003025	1.016
99003026	0.996

[1] See Attachment 2 for description of post stratum codes.

[FR Doc. 91-17202 Filed 7-16-91; 10:20 am]

BILLING CODE 3510-EA-M

federal register

Monday
July 22, 1991

Part IV

Department of Health and Human Services

Food and Drug Administration

21 CFR Parts 333 and 369

Topical Antimicrobial Drug Products for Over-the-Counter Human Use; Tentative Final Monograph for First Aid Antiseptic Drug Products; Proposed Rule

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration

21 CFR Parts 333 and 369

[Docket No. 75N-183F]

RIN 0905-AA06

Topical Antimicrobial Drug Products for Over-the-Counter Human Use; Tentative Final Monograph for First Aid Antiseptic Drug Products

AGENCY: Food and Drug Administration, HHS.

ACTION: Notice of proposed rulemaking.

SUMMARY: The Food and Drug Administration (FDA) is issuing a notice of proposed rulemaking in the form of an amended tentative final monograph that would establish conditions under which over-the-counter (OTC) first aid antiseptic drug products are generally recognized as safe and effective and not misbranded. FDA is issuing this notice of proposed rulemaking to amend the previous notice of proposed rulemaking on topical antimicrobial drug products after considering that rulemaking and public comments on it. (See the Federal Register of January 6, 1978, 43 FR 1210.) This proposal is part of the ongoing review of OTC drug products conducted by FDA.

DATES: Written comments, objections, or requests for oral hearing on the proposed regulation before the Commissioner of Food and Drugs by January 21, 1992. Because of the length and complexity of this proposed regulation, the agency is allowing a period of 180 days for comments and objections instead of the normal 60 days. New data by July 22, 1992. Comments on the new data by September 22, 1992. Written comments on the agency's economic impact determination by January 21, 1992.

ADDRESSES: Written comments, objections, new data, or requests for oral hearing to the Dockets Management Branch (HFA-305), Food and Drug Administration, rm. 1-23, 12420 Parklawn Dr., Rockville, MD 20857.

FOR FURTHER INFORMATION CONTACT: William E. Gilbertson, Center for Drug Evaluation and Research (HFD-210), Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857, 301-295-8000.

SUPPLEMENTARY INFORMATION: In the Federal Register of September 13, 1974 (39 FR 33103), FDA published, under § 330.10(a)(6) (21 CFR 330.10(a)(6)), an advance notice of proposed rulemaking to establish a monograph for OTC topical antimicrobial drug products, together with the recommendations of the Advisory Review Panel on OTC Topical Antimicrobial I Drug Products (Antimicrobial I Panel), which was the advisory review panel responsible for evaluating data on the active ingredients in this drug class. Interested persons were invited to submit comments by November 12, 1974. Reply comments in response to comments filed in the initial comment period could be submitted by December 12, 1974. In response to numerous requests, the agency issued a notice in the Federal Register of October 17, 1974 (39 FR 37066) granting an extension of the deadline for comments until December 12, 1974, and for reply comments until January 13, 1975.

In the Federal Register of January 6, 1978 (43 FR 1210), FDA published, under § 330.10(a) (7), a notice of proposed rulemaking to establish a monograph for OTC topical antimicrobial drug products, based on the recommendations of the Antimicrobial I Panel and the agency's response to comments submitted following publication of the advance notice of proposed rulemaking.

Interested persons were invited to submit objections or requests for oral hearing by February 6, 1978. In response to numerous requests to extend the time period for submitting objections or requests for oral hearing, the agency issued a notice in the Federal Register of February 3, 1978 (43 FR 4637) granting an extension of the deadline to March 6, 1978.

During this time period, the agency received 6 petitions that requested reopening the administrative record and 11 requests for an oral hearing. In a notice published in the Federal Register of March 9, 1979 (44 FR 13041), the agency deferred action on the requests for a hearing, but granted the petitions to reopen the record to allow interested persons to submit comments and any new or additional data by June 7, 1979, and reply comments by July 9, 1979. FDA also stated its intent to publish an updated (amended) tentative final monograph based on the review and evaluation of new submissions and a reevaluation of existing data.

In a notice published in the Federal Register of October 26, 1979 (44 FR 61609), the agency again reopened the administrative record for the submission of new data by March 26, 1980, and for comments on the new data by May 27, 1980. This action was taken to permit manufacturers to submit the results of testing to FDA as expeditiously as possible prior to establishment of a final monograph.

Subsequent to the June 7, 1979 closing date for the submission of new data, and prior to the October 26, 1979 reopening of the administrative record, data and information were submitted to FDA. In a notice published in the Federal Register of March 21, 1980 (45 FR 18398), the agency advised that it had reopened the administrative record for OTC topical antimicrobial drug products to allow for consideration of data and information that had been filed in the Dockets Management Branch after the administrative record on the tentative final monograph had officially closed on March 6, 1978. The agency concluded that any new data and information filed prior to March 21, 1980 should be available to the agency in developing a proposed regulation in the form of a tentative final monograph.

In a notice published in the Federal Register on January 5, 1982 (47 FR 436), the agency advised that it had again reopened the administrative record for OTC topical antimicrobial drug products to allow for consideration of the recommendations of the Advisory Review Panel on OTC Miscellaneous External Drug Products (Miscellaneous External Panel) on mercury-containing drug products. Interested persons were invited to submit comments by April 5, 1982, and reply comments by May 5, 1982. FDA stated that the proceeding to develop a monograph for mercury-containing drug products would be merged with the general proceeding to establish a monograph for OTC topical antimicrobial drug products.

In a notice published in the Federal Register on May 21, 1982 (47 FR 22324), the agency advised that it had again reopened the administrative record for OTC topical antimicrobial drug products to allow for consideration of the recommendations of the Miscellaneous External Panel on alcohol drug products. Interested persons were invited to submit comments by August 19, 1982, and reply comments by September 20, 1982. The notice stated that the proceeding to develop a monograph for alcohol drug products would be merged with the general proceeding to establish a monograph for OTC topical antimicrobial drug products.

In the Federal Register of September 7, 1982 (47 FR 39406), FDA issued a notice to reopen the administrative record for OTC topical antimicrobial drug products to allow for consideration of the Miscellaneous External Panel's recommendations on topical antimicrobial drug products used for the treatment of diaper rash. The agency discussed topical antimicrobial active

ingredients for this use in the Federal Register of June 20, 1990 (55 FR 25246).

In accordance with § 330.10(a)(10), the data and information considered by the Panels were put on public display in the Dockets Management Branch (address above), after deletion of a small amount of trade secret information. In response to the previous tentative final monograph and the advance notice of proposed rulemaking for mercury-containing drug products and the advance notice of proposed rulemaking for alcohol drug products, 4 drug manufacturers' associations, 44 drug manufacturers, 1 medical device manufacturer, 1 drug distributor, 2 medical schools, 2 research laboratories, 1 law firm, and 1 consulting firm submitted comments. Copies of the comments received are also on public display in the Dockets Management Branch (address above).

The advance notice of proposed rulemaking, which was published in the Federal Register of September 13, 1974 (39 FR 33103), was designated as a "proposed monograph" in order to conform to terminology used in the OTC drug review regulations (21 CFR 330.10). Similarly, the notice of proposed rulemaking, which was published in the Federal Register of January 6, 1978 (43 FR 1210), was designated as a "tentative final monograph." The present document is also designated as a "tentative final monograph." The legal status of each tentative final monograph, however, is that of a proposed rule.

This antimicrobial rulemaking is broad in scope, encompassing products that may contain the same active ingredients, but are labeled and marketed for different intended uses. For example, one group of products is primarily used by consumers for "first aid" and includes skin antiseptics, skin wound cleansers, and skin wound protectants. Another group of products is used by consumers on a more frequent, even daily basis, and includes products for personal use in the home, such as when caring for invalids and during family illness. Still a third group of products is generally intended for use by health professionals and includes health-care personnel handwashes, patient preoperative skin preparations, and surgical hand scrubs.

In order to expedite the completion of the first aid section of the antimicrobial monograph, the agency is publishing a separate tentative final monograph for these products. The non-first aid uses of topical antimicrobials will be addressed in a future issue of the Federal Register. Although the amended tentative final monographs for first-aid antiseptics and non-first aid uses of topical antimicrobials are being published separately, both categories will eventually be included under part 333 (21 CFR part 333).

The agency also has decided that OTC topical antimicrobial and topical antibiotic drug products should be included within the same monograph. Although an advance notice of proposed rulemaking to establish a monograph for OTC topical antibiotic drug products was published under part 342 (21 CFR part 342) on April 1, 1977 (42 FR 17642), the final monograph for those products was issued on December 11, 1987 (52 FR 47312) as a new subpart of the OTC topical antimicrobial monograph, 21 CFR part 333 subpart B—First Aid Antibiotic Drug Products.

Subpart A will cover first aid antiseptic drug products; subpart C will cover antifungal drug products; subpart D will cover acne drug products; and subpart E will cover non-first aid uses of topical antimicrobial drug products.

In this tentative final monograph (proposed rule) to establish subpart A of part 333 (21 CFR part 333), FDA states its position on the establishment of a monograph for OTC first aid antiseptic drug products only. This document addresses only those comments and data concerning the previous antimicrobial tentative final monograph that are related to "first aid uses." The agency will address all other submitted information at a later date.

This proposal constitutes FDA's reevaluation of the January 6, 1978 tentative final monograph based on the comments received and the agency's independent evaluation of the Miscellaneous External Panel's reports on OTC alcohol and mercury-containing drug products and the comments received. The following sections of the January 6, 1978 tentative final monograph for topical antimicrobial drug products are being addressed in this document: §§ 333.1, 333.3, 333.20, 333.40, 333.45, 333.65, 333.80, 333.90, 333.92, and 333.93. The following sections of the advance notice of proposed rulemaking for alcohol drug products are being addressed in this document: §§ 333.55 and 333.98. Modifications have been made for clarity and regulatory accuracy and to reflect new information. Such new information has been placed on file in the Dockets Management Branch (address above). These modifications are reflected in the following summary of the comments and FDA's responses to them. (See Part I.)

The OTC drug procedural regulations (21 CFR 330.10) now provide that any testing necessary to resolve the safety or effectiveness issues that formerly resulted in a Category III classification, and submission to FDA of the results of that testing or any other data, must be done during the OTC drug rulemaking process before the establishment of a final monograph. Accordingly, FDA will no longer use the terms "Category I" (generally recognized as safe and effective and not misbranded), "Category II" (not generally recognized as safe and effective or misbranded), and "Category III" (available data are insufficient to classify as safe and effective, and further testing is required) at the final monograph stage, but will use instead the terms "monograph conditions" (old Category I) and "nonmonograph conditions" (old Categories II and III). This document retains the concepts of Categories I, II, and III at the tentative final monograph stage.

The agency advises that the conditions under which the drug products that are subject to this monograph would be generally recognized as safe and effective and not misbranded (monograph conditions) will be effective 12 months after the date of publication of the final monograph in the Federal Register. On or after that date, no OTC drug product that is subject to the monograph and that contains a nonmonograph condition, i.e., a condition that would cause the drug to be not generally recognized as safe and effective or to be misbranded, may be initially introduced or initially delivered for introduction into interstate commerce unless it is the subject of an approved application. Further, any OTC drug product subject to this monograph that is repackaged or relabeled after the effective date of the monograph must be in compliance with the monograph regardless of the date the product was initially introduced or initially delivered for introduction into interstate commerce. Manufacturers are encouraged to comply voluntarily with the monograph at the earliest possible date.

In the advance notice of proposed rulemaking for OTC topical antimicrobial drug products (39 FR 33103), the agency suggested that the conditions included in the monograph (Category I) be effective 30 days after the date of publication of the final monograph in the Federal Register and that the conditions excluded from the monograph (Category II) be eliminated from OTC drug products effective 6 months after the date of publication of the final monograph, regardless of

whether further testing was undertaken to justify their future use. Experience has shown that relabeling of products covered by the monograph is necessary in order for manufacturers to comply with the monograph. New labels containing the monograph labeling have to be written, ordered, received, and incorporated into the manufacturing process. The agency has determined that it is impractical to expect new labeling to be in effect 30 days after the date of publication of the final monograph. Experience has shown also that if the deadline for relabeling is too short, the agency is burdened with extension requests and related paperwork.

In addition, some products will have to be reformulated to comply with the monograph. Reformulation often involves the need to do stability testing on the new product. An accelerated aging process may be used to test a new formulation; however, if the stability testing is not successful, and if further reformulation is required, there could be a further delay in having a new product available for manufacture. The agency wishes to establish a reasonable period of time for relabeling and reformulation in order to avoid an unnecessary disruption of the marketplace that could not only result in economic loss, but also interfere with consumers' access to these drug products. Therefore, the agency is proposing that the final monograph be effective 12 months after the date of its publication in the **Federal Register**. The agency believes that within 12 months after the date of publication most manufacturers can order new labeling and reformulate their products and have them in compliance in the marketplace. If the agency determines that any labeling for a condition included in the final monograph should be implemented sooner than the 12-month effective date, a shorter deadline may be established. Similarly, if a safety problem is identified for a particular nonmonograph condition, a shorter deadline may be set for removal of that condition from OTC drug products.

All "OTC Volumes" cited throughout this document refer to the submissions made by interested persons pursuant to the call-for-data notice published in the **Federal Register** of January 7, 1972 (37 FR 235) or to additional information that has come to the agency's attention since publication of the advance notice of proposed rulemaking. The volumes are on public display in the Dockets Management Branch (address above).

I. The Agency's Tentative Conclusions on the Comments and Reply Comments

A. General Comments

1. Two comments contended that OTC drug monographs are interpretive, as opposed to substantive, regulations. One comment referred to statements on this issue submitted earlier to other OTC rulemaking proceedings.

The agency addressed this issue in paragraphs 85 through 91 of the preamble to the procedures for classification of OTC drug products, published in the **Federal Register** of May 11, 1972 (37 FR 9464 at 9471 to 9472), and in paragraph 3 of the preamble to the tentative final monograph for OTC antacid drug products, published in the **Federal Register** of November 12, 1973 (38 FR 31260). FDA reaffirms the conclusions stated in those documents. Court decisions have confirmed the agency's authority to issue substantive regulations by informal rulemaking. (See, e.g., *National Nutritional Foods Association* v. *Weinberger*, 512 F.2d 688, 696–698 (2d Cir. 1975) and *National Association of Pharmaceutical Manufacturers* v. *FDA*, 487 F. Supp. 412 (S.D.N.Y. 1980), *aff'd*, 637 F.2d 887 (2d Cir. 1981).)

2. Two comments expressed concern over the amount of time that would be allowed for the relabeling of products after publication of the final monograph, citing the "Statement of Inflation Impact Potential" for the tentative final monograph which allowed a period of 7 to 12 months for manufacturers to implement labeling changes. One of the comments stated that the final monograph should allow at least 12 months to implement any required labeling changes. The other comment stated that such a period would be adequate for most regular production items, but would place a hardship on manufacturers with respect to infrequently produced products (e.g., once a year) and slow-moving items. This comment suggested an approach that would require all new labels ordered to comply in 6 months, all labels placed on products to comply in 18 months, and labels on all products shipped to comply in 24 months. The comment stated that this approach would allow labeling inventories for infrequently produced and slow-moving items to be depleted and would accommodate the agency's objectives and minimize the cost burden imposed on manufacturers and ultimately on consumers.

The agency agrees that a reasonable period of time should be provided for relabeling. As discussed more fully in the preamble of this document, the agency is proposing to extend this period so that the final monograph will be effective 12 months after the date of its publication in the **Federal Register**. The agency believes that, within this time, most manufacturers can have their products, including those infrequently produced, in compliance with the final monograph.

3. One comment expressed concern that scientific interpretations of testing data may differ between pharmaceutical manufacturers and FDA staff. The comment requested that the OTC drug review procedures provide an opportunity for a hearing prior to a final decision on a petition to reclassify an OTC drug product from Category III to Category I when genuine factual or scientific issues are raised concerning a drug's conformity with an OTC drug monograph.

This comment was submitted before the decision in *Cutler* v. *Kennedy*, 475 F. Supp. 838 (D.D.C. 1979). Before this decision was rendered, the OTC drug review procedural regulations in 21 CFR 333.10 allowed the continued marketing and testing of a Category III condition after a final monograph had been issued. Because of the court decision in *Cutler* v. *Kennedy*, the agency revised the OTC drug review procedural regulations so that all Category III testing must be completed prior to publication of a final monograph, if a manufacturer wants to upgrade a condition to Category I before the establishment of a final monograph. (See "Revision of Procedures Relating to Category III; Final Rule," published in the **Federal Register** of September 29, 1981, 46 FR 47730.)

Along with the publication of these revised procedures, the agency published a policy statement that provides for an exchange of information, including agency "feedback," on Category III test data between the agency and pharmaceutical manufacturers prior to publication of a final monograph. (See "Over-the-Counter (OTC) Drug Review Policy Statement," published in the **Federal Register** of September 29, 1981, 46 FR 47740.) The agency acknowledges that scientific interpretations of testing data may differ and believes that this "feedback" policy affords an adequate mechanism for pharmaceutical manufacturers and FDA to discuss air interpretations of testing data prior to a final monograph. In addition, under § 330.10(a)(7) interested parties may request an oral hearing after publication of a tentative final monograph. The agency believes that the existing regulations and the new "feedback"

policy provide adequate opportunities for pharmaceutical manufacturers to discuss data interpretations with FDA.

4. One comment stated that the agency should initiate revocation of new drug applications (NDA's) for products covered by the antimicrobial monograph upon publication of the final monograph. The comment contended that this would end continued use of claims that were approved under the NDA but are prohibited by the monograph, thus avoiding inequities in the industry and confusion in the marketplace.

The agency agrees with the comment that inequities and confusion should be avoided. After a final rule for OTC first aid antiseptic drug products is published, but before it becomes effective, the agency intends to publish in the Federal Register a notice of opportunity for hearing on a proposal to withdraw approval of new drug applications for products within the scope of the final monograph for OTC antimicrobial drug products.

5. One comment requested that Category III drugs be placed in Category I because they have already been extensively tested and have long been proven in the marketplace. According to the comment, if manufacturers consider it economically unfeasible to conduct the extensive Category III tests (43 FR 1210 at 1239 to 1245) the public would subsequently be deprived of drugs that it has found beneficial for self-medication for many decades. The comment stated that any currently marketed OTC drug that may later be proven unsafe, or whose claimed indications may be shown to be unwarranted, may be properly placed in Category II. However, the comment concluded that OTC drugs for which the Panel or the agency is merely seeking additional data should not be deleted from Category I while such data are being sought.

If the agency has classified an ingredient in Category III, it is because the available data are insufficient to classify the ingredient as generally recognized as safe and effective. Such ingredients cannot appropriately be put in Category I unless sufficient additional data are submitted to the rulemaking. This comment was submitted before the decision in *Cutler* v. *Kennedy*, 475 F. Supp. 838 (D.D.C. 1979), in which the OTC drug review procedural regulations in 21 CFR 333.10 that allowed the continued marketing and testing of a Category III condition after a final monograph were declared invalid. As stated in comment 3, because of this court decision, the agency has revised the OTC drug review procedural regulations so that all Category III testing must be completed prior to publication of a final monograph. Thus, it is not possible for the agency to affirmatively permit Category III drugs, which will be considered nonmonograph conditions, to remain on the market after the final monograph becomes effective, even if additional testing is being conducted to obtain data to support a Category I or monograph classification.

6. One comment stated that "removal from the marketplace of products which have been placed in Category II such as iodine, and the failure to include in the monograph various substances that have proven themselves in the marketplace for many years, will inevitably require the public to resort to more expensive but unnecessary substitutes."

The agency is proposing that several OTC topical antimicrobial ingredients, which have been in the marketplace for many years, be reclassified as Category I in this tentative final monograph under the new category "first aid antiseptic." Thus, these ingredients, including iodine, would not have to be removed from the marketplace. Previously marketed ingredients that have not been demonstrated to be safe and effective for any OTC use and that, therefore, are not included in any OTC drug monograph cannot legally be marketed without an approved application. The economic impact of this amended proposed rule on first aid antiseptic drug products is discussed elsewhere in this document.

7. One commenter pointed out that under "Subpart B—Active Ingredients" of the 1978 tentative final monograph, no CFR part number was assigned to the category "skin antiseptic." However, part numbers were assigned to other categories without any Category I ingredients, with the term "reserved" in parentheses. The comment requested that this omission be corrected in the amended tentative final monograph.

The omission pointed out by the comment was an oversight. However, it is no longer necessary to assign a CFR part number to the category "skin antiseptic," because skin antiseptics have been included in the broader category identified in this tentative final monograph as first aid antiseptics. (See comment 13.) All Category I first aid antiseptic active ingredients have been listed in the amended tentative final monograph under subpart A, § 333.10.

8. One comment submitted the final report of a 24-month study on the chronic toxicity of triclocarban as a petition to reopen the administrative record. Several comments had previously requested an extension of time from the March 26, 1980 deadline for the closing of the administrative record for submission of new data on conditions classified in Category III in the tentative final order, stating that the submission of the final report on the ongoing 2-year triclocarban toxicity study would not be completed by this deadline. The comments requested that the deadline for submission of new data be extended until the submission of this final report or, in the alternative, that FDA assure that the final report would be accepted and considered in this amended tentative final monograph.

In the notice of proposed rulemaking (43 FR 1210 at 1233), the agency requested that a 24-month study on the chronic toxicity of triclocarban be repeated. In response to this request, another 24-month study was initiated promptly, but because of the 2-year duration of the study, the final report was not submitted to the agency until May 27, 1981. To make this amended notice of proposed rulemaking as complete as possible, the agency has included the final report of the study in the administrative record and has considered the results of this study elsewhere in this document. (See comment 47.) Thus, the comment's request to extend the deadline for the submission of new data relating to triclocarban has been granted.

B. General Comments on Antimicrobials

9. Several comments objected to some of the specific statements of identity, e.g., "skin wound cleanser," "skin wound protectant," and "skin antiseptic." One comment stated that the word "skin" was superfluous because all OTC antiseptics are intended only for use on the skin. Another comment contended that the statement of identity "antiseptic" is preferable to "skin antiseptic" because these products are used on cuts, scratches, and mucous membranes as well as skin. One comment questioned whether consumers understand the statement of identity "skin wound protectant" and recommended that FDA adopt more familiar terminology, such as "first aid product." Other comments requested that Category I skin wound cleansers or skin wound protectants that contain antimicrobial ingredients be allowed a statement of identity that recognizes their antimicrobial activity, such as "first aid skin antiseptic," "minor antiseptic," "mild antiseptic," or "antimicrobial skin wound cleanser."

Based upon the comments, the agency believes that more familiar terminology could be used as the statement of identity and that the word "skin" should

not be required in the statement of identity for these products. In reviewing the indications recommended by the Panel for skin antiseptics, skin wound protectants, and skin wound cleansers, the agency identified the phrase "first aid product" as common to these drug categories. "First aid" is also a term that is frequently included in the labeling of OTC topical antiseptic drug products, reflects the intended OTC use of these products, is more familiar terminology to consumers, and is readily understood by consumers. Therefore, the agency is proposing the term "first aid antiseptic" as the statement of identity for OTC topical antimicrobial active ingredients included in this tentative final monograph. The agency has no objection to the statement "first aid antiseptic for the skin" or "first aid skin antiseptic" appearing elsewhere in the labeling of these products as additional information to the consumer, provided it does not appear in any portion of the labeling required by the monograph and does not detract from such required information.

10. Several comments argued that antimicrobial soap products making cosmetic claims only are not subject to regulation as OTC drugs and should not be considered in a review of drug effectiveness. The comments contended that if the intended use of antimicrobial soaps is stated solely in terms of deodorant effect, these products are not properly subject to regulation as OTC drugs. In addition, the comments stated that the OTC drug labeling requirements for antimicrobial soaps are unduly restrictive and uninformative. One comment pointed out that prior FDA regulations have recognized that personal cleanliness products (including both soaps and detergents) and underarm deodorants are cosmetic products, citing 21 CFR 720.49(c) (10) and FDA Trade Correspondence TC–26, February 9, 1940.

Some comments objected to the requirement that microbial reduction be established to demonstrate the deodorant effectiveness of OTC antimicrobial soaps because a direct correlation between bacterial reduction and the reduction of body odor has not been scientifically determined. One comment cited three studies (Refs. 1, 2, and 3) to support this contention. The comments requested that the final antimicrobial monograph apply only to antimicrobial soaps that make specific drug claims that any reference to deodorant claims be deleted from the monograph.

Other comments requested that the labeling for antimicrobial soaps be expanded to give more emphasis to the deodorant activity of these products. The comments objected to the limitation of phrases and the restrictions on the use of the phrases "reduces odor" and "deodorant soap" as well as to phraseology concerning deodorant usage in § 333.80 of the monograph.

Several comments objected to the proposed indication "antimicrobial soap" (§ 333.80(b) (1)) and requested that it either be deleted or modified to include deodorant claims. The comments contended that it is redundant and serves no purpose to require that the label of an antimicrobial soap contain the statement "antimicrobial soap" both as an indication and as a statement of identity (§ 333.80(a)).

One comment stated that this labeling requirement represents a misuse of the word "indications" because the permitted terms "antimicrobial" or "antibacterial" do not inform consumers of the intended use of the product in terms likely to be understood by the ordinary individual. The comments stated that because these labeling requirements do not adequately convey to consumers that the principal use and benefit to be derived from the use of antimicrobial soaps is the deodorant effect, these labeling requirements may not only confuse consumers but also may deny them truthful and useful information about these products.

The agency has carefully reevaluated this issue and clarifies that the OTC drug monographs promulgated under 21 CFR part 330 cover drug ingredients and indications, not cosmetic claims. The Federal Food, Drug, and Cosmetic Act (the act) principally defines a "drug" as an article "intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease" or "intended to affect the structure or any function of the body * * *." (21 U.S.C. 321(g)(1)). The act defines a "cosmetic" as an article "intended to be rubbed, poured, sprinkled, or sprayed on, introduced into, or otherwise applied to the human body or any part thereof for cleansing, beautifying, promoting attractiveness, or altering the appearance * * *." (21 U.S.C. 321(i)(1)). The intended use of the product, therefore, determines whether the product is a "drug," a "cosmetic," or both. This intended use may be inferred from the product's labeling, promotional material, advertising, and any other relevant factor. (See, e.g., *National Nutritional Foods Association* v. *Mathews*, 557 F.2d 325, 334 (2d Cir. 1977). A manufacturer's subjective claims of intent may be pierced to find its actual intent on the basis of objective evidence. *National Nutritional Foods Association* v. *FDA*, 504 F.2d 761, 789 (2d Cir. 1974).

The agency emphasizes that the previous tentative final monograph and this amended proposal cover only those antimicrobial products that are drugs or are both drugs and cosmetics, and are not applicable to the wide variety of products that are only cosmetics but that contain antimicrobial ingredients. The agency notes that most currently marketed antimicrobial bar soaps are not viewed by the consuming public as drugs but as products providing personal cleaning and deodorizing benefits. The agency agrees that separate regulations are required to govern the safety of cosmetic products containing antimicrobial ingredients. In comment 12 of the 1978 tentative final monograph (43 FR 1210 at 1212) the agency stated its intention not to require NDA's for products containing a Category I antimicrobial ingredient at greater than preservative levels and that make no drug claims. This position remains unchanged. Therefore, the presence of an antimicrobial ingredient does not, in and of itself, make a product a drug provided that no drug claim is made. However, the antimicrobial ingredient included in a cosmetic product may not exceed the concentration provided for in an applicable monograph.

As the comments have pointed out, the agency has in the past acknowledged "deodorancy" to be a cosmetic claim. Soap products that contain antimicrobial ingredients will be considered "cosmetics," and not "drugs," if only deodorant claims (or other cosmetic claims) are made for the products. The agency has previously stated that the mere presence of an antimicrobial ingredient in a product labeled for deodorant use, with the ingredient identified only in the ingredient list and no reference to its antimicrobial properties stated elsewhere in the labeling, would not cause the product to be considered a drug (Ref. 4).

However, any broader claims that represent or suggest a drug use for the product would subject it to regulation as a drug. For example, the agency considers terms such as "antibacterial," "antimicrobial," or "kills germs" in the labeling of deodorant soap products to imply that the product will have a therapeutic effect. Such statements would constitute a drug claim for the product. Likewise, statements in the labeling of a deodorant soap product such as "antimicrobial for deodorization" or "kills germs that

cause body odor" will cause the product to be a drug. Further, the term "active ingredient(s)" used anywhere in labeling would imply that the product possesses a drug-like property and would also cause the product to be a drug.

In summary, deodorant effectiveness and related claims in the labeling of soap products that contain antimicrobial ingredients but make only cosmetic claims will not be considered further in this document. Accordingly, the agency is deleting previously proposed § 333.80. However, if a manufacturer elects to market such a product as a drug (e.g., by including labeling as an "antimicrobial"), the product is a drug and is required to demonstrate efficacy, even if the labeling claim is only for a deodorant effect. Testing guidelines for antimicrobial claims will be addressed in an amended tentative final monograph covering non-first aid topical antimicrobial indications, to be published in a future issue of the **Federal Register.**

In addition, the agency did not receive any data on the use of antimicrobial soaps specifically labeled for first aid use. Consequently, antimicrobial soaps are not being included in this tentative final monograph for this use. Other drug uses (e.g., for general health care) will also be addressed in a future issue of the **Federal Register.**

References

(1) Prince, H.N., and J.A. Rodgers, "Studies on the Aerobic Axillary Microflora Employing a Standardized Swabbing Technique (Total Counts, Speciation and Ecological Drift)," Cosmetics and Perfumery, 89:25–30, 1974.

(2) Dravnieks, A., et al., "Influence of an Antimicrobial Soap on Various Effluents From Axillae," Journal of the Society of Cosmetic Chemists, 19:611–626, 1968.

(3) Cowen, R.A., "Relative Merits of 'In Use' and Laboratory Methods for the Evaluation of Antimicrobial Products," Journal of the Society of Cosmetic Chemists, 25:307–323, 1974.

(4) Memorandum of Meeting between Armour Dial, Inc., and FDA, March 9, 1983, coded MM0001, Docket No. 75N–0183, Dockets Management Branch.

11. One comment requested that scrubbing devices, such as brushes or sponges, that are impregnated with approved antimicrobial ingredients be included in the monograph.

Although the comment intended to address professional antimicrobial uses, the question of impregnated scrubbing devices may also be relevant to first aid uses. This amended tentative final monograph does not specifically provide for the use of devices such as brushes or sponges impregnated with antimicrobials. These devices are not included in the monograph because the monograph is intended to regulate OTC drug active ingredients, not device delivery systems, except to the extent that the method of application is important to the OTC drug's safety or effectiveness, and the device employed is legally available. Under such circumstances, the monograph may specify the use of the device for the specific drug.

The agency does not believe that it is necessary to include specific references to brush or sponge delivery systems in the first aid antiseptic monograph. If a topical antimicrobial active ingredient is used to impregnate a scrubbing device such as a brush or sponge as the method of application of the drug, the topical antimicrobial component continues to be regulated as a drug (and must conform to the applicable conditions of the final monograph if the ingredient is included in the monograph for the product's labeled indications), and the instrument must satisfy the device requirements under the act. For example, a brush impregnated with an antimicrobial active ingredient and intended for use as a first aid antiseptic must conform to the first aid antiseptic requirements included in Subpart A of this proposed monograph as well as any appropriate device requirements.

12. One comment expressed concern that the tentative final monograph failed to provide consumers with an antibacterial skin cleanser for home use. The comment noted that, in addition to professional health care personnel, many consumers have a need for cleansing products containing antibacterial agents for the purpose of promoting good individual and family hygiene. Potential uses cited for such products included: (1) To reduce bacteria on the hands and face to a greater extent than can be accomplished with ordinary soap, and to prevent accumulation of bacteria from potential sources of contamination. The following examples were cited: Cleansing oneself after changing a baby's diaper, or after assisting aged or ill members of the household with their toilet needs, and before preparing a family meal. (2) The added benefit of an antibacterial cleanser for the minute cuts and abrasions from shaving and other minor traumas. (3) The need for an antibacterial cleanser other than bar soap on local parts of the body, such as the face, because soap (alkali salts of fatty acids) can be irritating or too drying for some individuals' needs. The comment recommended a new product class under proposed § 333.90(a) (skin antiseptic) to be identified as "Antimicrobial (or Antibacterial) Personal Cleanser" with claims such as "decreases bacteria on the skin" and "contains an antibacterial agent." The comment also suggested that the 10-day maximum use limitation would not be appropriate for this product class, but use could be restricted to 5 or 10 times daily.

The agency believes that the comment's recommendation has merit; however, this document is limited in scope to first aid antiseptic drug products. The agency will address the issue of cleansing products containing antibacterial agents for the purpose of promoting good individual and family health care in the non-first aid uses segment of the amended tentative final monograph, in a future issue of the **Federal Register.**

C. Comments on Definitions

13. Several comments objected to the definition of "skin antiseptic" in proposed § 333.3(f): "A nonirritating, antimicrobial-containing preparation that prevents overt skin infection." The comments asserted that this definition requires total effectiveness (that is, antimicrobial activity against all infective agents), that this is an unreasonable and unrealistic definition, and that, at present, no testing methods conclusively demonstrate total effectiveness. The comments stated that the proposed definition is too restrictive and cited three definitions of an antiseptic that do not include the concept of prevention of infection (Refs. 1, 2, and 3). In addition, the comments pointed out that the statutory definition of an antiseptic (section 201(o) of the act) is not subject to the discretionary enlargement that was recommended by the agency in the tentative final monograph (43 FR 1210 at 1215). The comments submitted alternative definitions and requested that one of them be adopted. Two comments recommended the following definition of skin antiseptic: "A nonirritating antimicrobial-containing preparation that kills or inhibits the growth of microorganisms on the skin."

As discussed earlier in this document, the agency is proposing that skin antiseptics, as well as skin wound protectants and skin wound cleansers, be included in one category called "first aid antiseptics." Thus, a separate definition of "skin antiseptic" is no longer necessary, and § 333.3(f) of the previous tentative final monograph is not being included in this amended tentative final monograph. It is generally recognized that the chief purpose of a first aid antiseptic is to kill or prevent the growth of bacteria that may cause

infection. Therefore, the agency is proposing in this amended tentative final monograph to define the term "first aid antiseptic" as follows: "An antiseptic-containing drug product applied topically to the skin to help prevent infection in minor cuts, scrapes, and burns." The agency believes this definition is consistent with section 201(o) of the act and is more realistic than the previously proposed definition because it does not require total effectiveness against all infective agents, the concern expressed by the comments.

Regarding testing, it should be noted that the Panel expressed concern over the confusion concerning the definition and use of the term "antiseptic." The Panel believed that the definition of antiseptic had been interpreted as activity against infection or microbial sepsis (39 FR 33103 at 33114). The term "antiseptic" is comparable to accepted definitions for a disinfectant. The Panel attempted to eliminate the confusion "by developing a rigorous definition of a skin antiseptic" (39 FR 33114). The Panel stated that claims stating or implying an effect against microorganisms must be supported by controlled human studies demonstrating prevention of infection. The agency indicated in the tentative final monograph (43 FR 1210 at 1211) that the testing regimens were not intended to be more burdensome than needed to prove safety and effectiveness, as required by law. However, neither the Panel nor the agency proposed a specific protocol to test claimed "skin antiseptic" products. In the tentative final monograph, the agency proposed that the testing guidelines for products intended for use by health professionals be used (43 FR 1216). The agency continues to believe that products that meet these requirements are acceptable as "first-aid antiseptics," but it is not necessary for first aid antiseptics to meet these more rigid testing requirements for products intended for use by health professionals.

In this document, the agency is proposing a more consumer-oriented indication for first aid antiseptics than the indications previously proposed in § 333.90 for skin antiseptics. The new indication is as follows: "First aid to help" [select one of the following: "prevent," ("decrease" ("the risk of" or "the chance of")), ("reduce" ("the risk of" or "the chance of")), "guard against," or "protect against"] [select one of the following: "infection," "bacterial contamination," or "skin infection"] "in minor cuts, scrapes, and burns." Manufacturers choosing to market a first aid product with this claim need only meet the requirements specified in the proposed monograph. To assist manufacturers in meeting these requirements, the agency is also providing procedures for testing a "first aid antiseptic." (See comment 56.)

References

(1) "Webster's New Collegiate Dictionary," G. and C. Merriam Co., Springfield, MA, 1975, s.v. "antiseptic."

(2) "Dorland's Illustrated Medical Dictionary," 24th Ed., W.B. Saunders Co., Philadelphia and London, 1965, s.v. "antiseptic."

(3) Harvey, S.C., "Antiseptics and Disinfectants; Fungicides; Ectoparasiticides," in "The Pharmacological Basis of Therapeutics," 5th Ed., edited by L.S. Goodman and A. Gilman, The Macmillan Co., New York, p. 988, 1975.

14. Several comments objected to the definitions of the proposed skin wound cleanser and skin wound protectant categories. One comment stated that although the Panel was only charged to evaluate antimicrobial ingredients, its recommendations for the skin wound protectants and skin wound cleansers clearly extended to nonantimicrobial ingredients. This comment recommended that FDA modify the antimicrobial monograph to make it clear that it is limited to products with antimicrobial ingredients. Two comments objected to the revision in the definition of "Skin Wound Protectant" in § 333.3(h) of the tentative final monograph, which states in part "* * * it provides a protective physical barrier and a chemical (antimicrobial) barrier * * *." The comments contended that the Panel's definition of skin wound protectant in § 333.3(f) should be adopted: "A safe, non-irritating preparation applied to small cleansed wounds which provides a protective (physical and/or chemical) barrier and neither delays healing nor favors the of micro-organisms."

One comment requested that FDA include recommendations on the safety and effectiveness of the nonantimicrobial ingredients that act as physical barriers in skin wound protectants. Another comment submitted data on a cream physical barrier product without a claimed active antimicrobial agent to show that the product is safe and nonirritating, provides a protective barrier, does not delay wound healing or favor the growth of microorganisms, and therefore meets all of the criteria for a skin wound protectant as defined by the Panel (Ref. 1). This comment argued that the addition of an antimicrobial ingredient cannot contribute to the claimed effectiveness of this product when all of the efficacy criteria have been met without it. The comment concluded that FDA should either return to the Panel's definition, which does not require a chemical barrier, or modify the definition and testing required for a skin wound protectant in the previous tentative final monograph in such a way that the antimicrobial ingredient will contribute to the claimed efficacy of the product.

The agency agrees with the comment that contended that skin wound cleansers and skin wound protectants without active antimicrobial ingredients do not fall within the scope of the antimicrobial rulemaking. This amended tentative final monograph applies to products containing antimicrobial ingredients for first aid antiseptic use. As discussed in comment 13, the definitions for skin wound cleanser and skin wound protectant are no longer included in this amended tentative final monograph. The agency will discuss the data submitted by the comment for a product containing no antimicrobial ingredient, but with protective claims, in the rulemaking for OTC skin protectant drug products, in a future issue of the **Federal Register**.

Reference

(1) Comment No. C00107, Docket No. 75N-0183, Dockets Management Branch.

D. Comments on Labeling

15. Several comments contended that FDA does not have the authority to restrict OTC labeling claims to exact wording, to the exclusion of what the comments described as other "equally truthful claims for the products." One comment pointed out that numerous other meaningful and truthful statements will provide useful information and will enhance the safe and effective use of these products. Several comments maintained that manufacturers have a constitutional right to use any truthful, nonmisleading labeling under the first amendment. To support their position, the comments cited *Bigelow* v. *Virginia*, 421 U.S. 809 (1975); *Virginia State Board of Pharmacy* v. *Virginia Citizens Consumer Council, Inc.*, 425 U.S. 748 (1976); *Linmark Associates, Inc.* v. *Willingboro*, 431 U.S. 85 (1977); *Bates* v. *State Bar of Arizona*, 433 U.S. 350 (1977); *Federal Trade Commission* v. *Beneficial Corp.*, 542 F.2d 611, 97 S. Ct. 1679 (1977); and *Warner-Lambert Co.* v. *Federal Trade Commission*, 562 F.2d 749 at 768 (D.C. Cir. 1977).

In the **Federal Register** of May 1, 1986 (51 FR 16258), the agency published a final rule changing its labeling policy for stating the indications for use of OTC

drug products. Under 21 CFR 330.1(c)(2), the label and labeling of OTC drug products are required to contain in a prominent and conspicuous location, either (1) the specific wording on indications for use established under an OTC drug monograph, which may appear within a boxed area designated "APPROVED USES"; (2) other wording describing such indications for use that meets the statutory prohibitions against false or misleading labeling, which shall neither appear within a boxed area nor be designated "APPROVED USES"; or (3) the approved monograph language on indications, which may appear within a boxed area designated "APPROVED USES," plus alternative language describing indications for use that is not false or misleading, which shall appear elsewhere in the labeling. All other OTC drug labeling required by a monograph or other regulation (e.g., statement of identity, warnings, and directions) must appear in the specific wording established under the OTC drug monograph or other regulation where exact language has been established and identified by quotation marks, e.g., 21 CFR 201.63 or 330.1(g).

In the previous tentative final monograph, supplemental language relating to indications had been proposed and captioned as "*Other allowable statements*" in §§ 333.90, 333.92, and 333.93. Under FDA's revised labeling policy (51 FR 16258), such statements are included at the tentative final stage as examples of other truthful and nonmisleading language that would be allowed elsewhere in the labeling. In accordance with the revised labeling policy, such statements would not be included in a final monograph.

In preparing this amended tentative final monograph, the agency has reevaluated these "other allowable statements" to determine whether they should be incorporated, wherever possible, as part of the indications developed under the monograph. The "*Other allowable statements*" proposed in the previous tentative final monograph that are covered by this amended tentative final monograph appeared in § 333.90(b)(2) for skin antiseptic, in § 333.92(b)(2) for skin wound cleanser, and in § 333.93(b)(2) for skin wound protectant. The statement "provides a protective physical (and chemical) barrier" proposed for a skin wound protectant has been deleted in this tentative final monograph because it does not fall within the scope of the antimicrobial rulemaking. (See comment 14.) Other previously proposed "*Other allowable statements*" are discussed in comment 16.

16. Two comments suggested that the following labeling claims would be appropriate for first aid antiseptics: "degerms," "kills germs," "kills bacteria," "bactericidal" (if applicable as, for example, for alcohol), "contains antimicrobial ingredients," "microbiocidal," "first aid product," and "reduces the risk of infection." One of the comments argued that the labeling claims "prevents overt infection" or "controls infection" should be permitted if appropriate additional studies are provided.

Other previously proposed "*Other allowable statements,*" i.e., "contains antibacterial ingredient(s)," "contains antimicrobial ingredient(s)," "does not delay wound healing," and "nonirritating" are similar to the claims suggested by the comments, and the agency is evaluating all of these statements concurrently. The OTC drug review program establishes conditions under which OTC drugs are generally recognized as safe and effective and not misbranded. One aspect of the program is to develop standards for certain parts of the labeling of OTC drug products. FDA has found that it is simply not practical—in terms of time, resources, and other considerations—to set standards for all labeling found on OTC drug products. Accordingly, OTC drug monographs directly address only those labeling items that are related in a significant way to the safe and effective use of covered products by lay persons. These labeling items are the product statement of identity; names of active ingredients; indications for use; directions for use; warnings against unsafe use, side effects, and adverse reactions; and claims concerning mechanism of drug action.

The agency finds that most of the terms suggested by the comments and previously proposed as "*Other allowable statements,*" while descriptive of the action of first aid antiseptic products, do not relate in a significant way to the safe and effective use of these products and, therefore, are outside the scope of the monograph.

However, the OTC drug review is also intended to ensure that OTC drug products are not misbranded. Therefore, the agency evaluates claims made on OTC drug product labels on a product-by-product basis, under section 502 of the act (21 U.S.C. 352), to determine whether those claims are false or misleading. Any claim that is outside the scope of the monograph, even though it is truthful and not misleading, may not appear in any portion of the labeling that is required by the monograph. Such a claim also may not detract from the required information. Therefore, the claims requested by the comments or previously proposed as "*Other allowable statements,*" except for those discussed below, may be included on the labeling of first aid antiseptic drug products provided that they are not intermixed with labeling established by the monograph and that they are not false or misleading.

The agency does not believe the average consumer would understand the word "overt" in the phrase "prevents overt infection." As for the phrase "controls infection," the agency believes it may mislead the consumer into assuming the product is intended for use in treating an existing infection. The agency is proposing "helps prevent infection" as a suitable alternative to the two phrases above.

The agency believes that claims such as "degerms," "degerming," "kills germs," "kills bacteria," "bactericidal," and "microbiocidal" could be potentially misleading to the average consumer if directly associated with the term "infection" that is included in the indication because the terms "kill" and "-cidal" may be interpreted to imply elimination of all bacteria on the skin when, in fact, topical antiseptics used on the skin only decrease the number of certain bacteria. However, the agency acknowledges that these terms are familiar to the average consumer and may be useful in describing the product's action or intended effect. Although these terms are not included in the monograph, they may be included in labeling that is not intermixed with monograph labeling as described above.

17. One comment requested that the following phrases (or their equivalent) be added to the monograph: Under proposed § 333.92(b)(1), "to clean and kill germs in superficial wounds," and under proposed § 333.92(b)(2), "contains a safe and effective germ-killing active ingredient." The comment also suggested that the indication "contains antimicrobial ingredient," in proposed § 333.92(b)(2) for skin wound cleansers, be expanded to provide a lay definition of "antimicrobial ingredient" because most consumers would not fully understand the meaning of the statement.

The skin wound cleanser category (proposed § 333.92) is not included in this amended antimicrobial tentative final monograph. As discussed in comment 13, all antimicrobial-containing products to be used on minor cuts, scrapes, and burns are now included in a single category, i.e., first aid antiseptic drug products.

The agency has considered the comment's request to include additional phrases to expand and clarify the meaning of "antimicrobial ingredient." The agency agrees with the comment that labeling should be more informative and has provided several optional statements in § 333.50 of the amended tentative final monograph. However, as discussed in comment 16, the agency believes that a number of terms, e.g., "kills germs," are descriptive but outside the scope of the OTC drug monograph. If such terms are included in labeling, they may not appear in any portion of the labeling required by the monograph and may not detract from such required information.

18. One comment from a manufacturer of a skin wound protectant requested that the claim "protects against * * * diaper rash" be added to the list of indications in proposed § 333.93(b)(1) for skin wound protectants. The comment stated that its product enjoys considerable use in the treatment of diaper rash, but that if an indication for diaper rash is not included in the monograph, the product could not be promoted for one of its primary uses.

As noted in comment 13, the skin wound protectant category is not included in the amended antimicrobial tentative final monograph. In the Federal Register of September 7, 1982 (47 FR 39436), the administrative record for skin protectant drug products was reopened to include the recommendations on diaper rash drug products of the Advisory Review Panel on OTC Miscellaneous External Drug Products (Miscellaneous External Panel) because that Panel concluded that the use of skin protectants may prevent skin irritation associated with diaper rash (47 FR 39436 at 39439).

In the tentative final monograph for OTC skin protectant diaper rash drug products, the claim "protects * * * diaper rash" was proposed as a monograph claim (June 20, 1990; 55 FR 25204 at 25232). In the tentative final monograph for OTC topical antimicrobial diaper rash drug products, no antimicrobial ingredients were proposed as Category I for the claim "helps protect against skin infection associated with diaper rash" (June 20, 1990; 55 FR 25246 at 25281). Final agency decisions will appear in the final monographs for OTC diaper rash drug products in a future issue of the Federal Register.

19. One comment recommended that antimicrobial soaps be allowed to make claims relating to general health care and personal hygiene similar to the claims allowed for health-care personnel handwashes. The comment stated that an antimicrobial soap will reduce bacteria or the transfer of potentially pathogenic organisms in the home, and therefore serves as a preventive health care aid in controlling diseases such as impetigo, pyoderma, and erythrasma. To inform consumers of such benefits, the comment suggested that the "other allowable statements" for antimicrobial soaps be expanded to include some of the labeling for health-care personnel handwashes in proposed § 333.85(b)(1).

The agency will address these uses of such products in a future Federal Register notice.

20. One comment objected to that part of the directions for use for skin wound protectants (§ 333.93(d)) that states, "After gentle washing with soap and water, * * *." The comment contended that in certain instances "gentle washing with soap and water" does not constitute acceptable medical practice, and requested that the wording should simply be "apply small amount directly to the affected area." Two comments objected to that part of the directions for use that states "May be applied 1 to 3 times daily." One comment stated that such a limitation of use should be based on the active ingredient(s). The comment recommended the following wording: "Labeling should also contain the recommended time interval (if any) between applications required to provide a protective (physical and chemical) barrier on the skin." The other comment pointed out that first aid products are intended only for single or a few applications. This comment contended that labeling that implies repeated use will be confusing to the consumer and suggested substituting labeling that does not assume repeated use.

The agency believes that first aid of small superficial wounds begins with adequate cleaning of the wound and, therefore, disagrees with the comment's suggestion to delete all references to cleaning the wound. However, because alkaline soap may not be appropriate for use on damaged tissue, the agency proposes to replace the phrase "after gentle washing with soap and water," with the phrase "clean the affected area."

Regarding the directions to use 1 to 3 times daily, such a direction is appropriate for these products, will discourage unlimited and repeated use, and yet will allow for limited applications as needed after a bath or after washing.

Therefore, the agency is proposing the following general directions for use for first aid antiseptics in § 333.50(d): "Clean the affected area. Apply a small amount of this product on the area 1 to 3 times daily. May be covered with a sterile bandage."

21. One comment requested that the portion of the directions for skin wound protectants in proposed § 333.93(d) that states "* * * cover with sterile gauze if desired" be deleted because covering a wound may retard healing in some cases. The comment submitted no data to support its request.

The agency agrees with the comment that it is not always desirable to cover a wound. However, rather than deleting any reference to covering a wound, the agency believes that consumers should be informed if precautions should be taken when covering a wound. For first aid antiseptics that do not require special labeling concerning bandages, the agency is proposing the directions for use stated in comment 20.

22. Objecting to the proposed warning "Do not bandage tightly" (§ 333.92(c)(4)), one comment stated that the warning does not make sense in terms of the way in which quaternary ammonium skin wound cleansers such as benzalkonium chloride are generally used. In place of the proposed warning, the comment recommended more explicit instructions for use, e.g., "Apply and let dry, before bandaging," and submitted data to support its position that occlusion of the wound with a bandage does not interfere with the safety and effectiveness of the drug (Ref. 1).

As discussed by the Panel (39 FR 33103 at 33132), quaternary ammonium compounds can be irritating to the skin, and the degree of irritation is dependent on concentration and/or occlusion. The Panel stated "There is little irritation potential with the use concentration." Nevertheless, the Panel stated that these compounds should not be covered with occlusive bandaging (39 FR 33116) and recommended the following warning: "Use of solution with occlusive dressing is not advisable." In paragraph 57 of the previous tentative final monograph (43 FR 1210 at 1219), the warning against "occlusive dressing" was revised to "Do not bandage tightly," and included in the warning for all skin wound cleansers. Upon further review of this warning, in the context of the newly proposed first aid category, the agency is proposing not to include a general warning statement, but instead to evaluate each individual ingredient to determine if there is a need for such a statement. The agency has reviewed the data on benzalkonium chloride submitted by the comment (Ref. 1) and determined that they show that occlusion of the wound with a bandage did not interfere with healing of the wound. Accordingly, the

agency concludes that the warning "do not bandage tightly," previously proposed in § 333.92(c)(4), is not necessary for this ingredient at the use concentrations provided for in the proposed monograph. Likewise, the alternate warning previously proposed in § 333.99 in the professional labeling section, i.e., "Do not use solution with occlusive dressing," is no longer being included in the tentative final monograph. The agency has also determined that these warning statements are not necessary for the other two quaternary ammonium compounds included in this monograph. Benzethonium chloride has been shown to be not irritating or sensitizing in two studies on children with diaper rash (Refs. 2 and 3). Methylbenzethonium chloride, a derivative of benzethonium chloride, has been used to prevent and treat skin irritations caused by contact with urine, feces, and perspiration, and has low toxicity and local sensitizing properties (Ref. 4).

The agency notes that first aid antiseptics containing quaternary ammonium compounds are usually applied as solutions or sprays, and agrees with the comment that more explicit directions for use relating to bandaging after applying the product would be useful to consumers. The agency is also aware that a number of other first aid antiseptic ingredients are marketed as solutions or sprays. The agency believes it is appropriate to let a solution or spray dry first before covering the area with a sterile bandage. Accordingly, the agency is incorporating this information in the directions section of this tentative final monograph.

References

(1) Unpublished Clinical Wound Healing Studies on Medi-Quik® submitted by Sterling Drug, Inc., Comment No. SUP013, Docket No. 75N–0183, Dockets Management Branch.

(a) Statistical Analysis of Data from Efficacy Study of Medi-Quik® as a Skin Wound Protectant in Humans.

(b) Studies on Medi-Quik® as a Wound Protectant.

(2) Susca, L.A., and B.G. Genting, "Treatment of Diaper Rash," New York State Journal of Medicine, 69:2858–2862, 1960.

(3) Christian, J.R., and F. Gonzalez, "Topical Treatment of Acute and Chronic Diaper Rash with Amino Acid Creme," Clinical Medicine, 80, 1961.

(4) Osol, A., and R. Pratt, "The United States Dispensatory," 27th Ed., J.B. Lippincott Co., Philadelphia, p. 186, 1973.

23. Two comments objected to the warning "This product is not for use on wild or domestic animal bites. If you have an animal bite, consult your physician immediately," that was proposed for skin antiseptics (§ 333.90(c)(2)), skin wound cleansers (§ 333.92(c)(2)), and skin wound protectants (333.93(c)(2)). One comment pointed out that skin antiseptics, skin wound cleansers, and skin wound protectants may be particularly useful for cleansing or for first aid treatment of wounds, including animal bites, when medical treatment is not immediately available.

Acknowledging that consumers should not rely solely on self-medication for animal bites, the comment suggested the following warning: "If you have an animal bite, consult your physician immediately." The other comment recommended deleting the warning for skin antiseptics in proposed § 333.90(c)(2), arguing that it is inappropriate because consumers know they cannot rely solely upon antiseptics to treat animal bites and that they should be examined by a physician. The comment further contended that including this type of warning in the labeling may cause consumers to view other important labeling statements on OTC drug products with skepticism.

The agency agrees that most consumers would know that a severe injury from an animal bite needs medical attention; however, consumers may not be as aware of the dangers of the superficial bites of small animals. Although bites from small wild or domestic animals, such as racoons or cats, may appear to be minor cuts, they can result in skin infections or possibly even in rabies. Consequently, the agency believes that an animal bite warning is necessary on OTC first aid antiseptic drug products to warn consumers against relying on self-medication for any animal bite. However, the agency believes that, rather than having a separate warning for animal bites, it is preferable to add the term "animal bites" to the warning that lists other conditions requiring medical attention. Therefore, the agency is proposing the following revised warning for first aid antiseptic drug products in § 333.50(c)(1) in this tentative final monograph: "In case of deep or puncture wounds, animal bites, or serious burns, consult a doctor."

24. One comment objected to the proposed warnings against the use of skin antiseptics, skin wound cleansers, and skin wound protectants for more than 10 days (§§ 333.90(c)(3), 333.92(c)(3), and 333.93(c)(3)). The comment pointed out that these products are not recommended for daily use and that a warning that implies repeated use will be confusing to consumers. The comment also pointed out a discrepancy between the wording in the second sentence of the warning for skin antiseptics which states: "If the infection worsens or persists, see your physician," and the equivalent warning for skin wound cleansers and skin wound protectants, which states: "If the condition worsens or persists, see your physician." The comment maintained that the warning for skin antiseptics is confusing because it assumes that an infection has occurred when, in fact, none may have occurred, but the wound nevertheless requires medical attention. The comment suggested that all three warnings be replaced by one warning as follows: "If condition does not improve in 10 days, see your physician."

Another comment stated that the warnings in proposed § 333.93 (c)(3) and (c)(5) convey the same message. The warning in § 333.93(c)(3) states, "Do not use this product for more than 10 days. If the conditions worsen or persist, see your physician." The warning in § 333.93(c)(5) states, "If itching, redness, irritation, swelling or pain develops and persists for more than 1 week or increases, it may be a sign of infection or allergy. Discontinue use at once and consult your physician." The comment requested that this warning be deleted.

A third comment requested that alcohol drug products also be labeled with a warning to consult a physician if the condition worsens or persists for more than 1 week.

As noted in comment 13, the three categories formerly identified as skin antiseptic, skin wound protectant, and skin wound cleanser have all been included in the first aid antiseptic category, and all drugs in this category will bear the same warnings.

The agency disagrees that the statement limiting the period of use implies that the product is recommended for repeated daily use. The purpose of a statement limiting use of a product is to alert the consumer to the period of time that is reasonable for self-treatment of a condition and to convey the message that a condition that persists beyond this period should be treated by a doctor.

The agency agrees with the comments that the warnings in § 333.93 (c)(3) and (c)(5) convey the same message and that the statement in § 333.90(c)(3) "If the infection worsens or persists, see your physician" implies an existing infection and may cause confusion about when a physician should be consulted. The proposed warning in § 333.93(c)(5) could confuse consumers because it states that the user should stop using the product if itching, redness, swelling or pain develops or increases. These are the same symptoms that often occur after minor skin injury, the condition for

which topical first aid products are indicated. Therefore, for clarity, §§ 333.90(c)(3), 333.92(c)(3), and 333.93 (c)(3) and (c)(5) have been combined and revised. The proposed warning, redesignated § 333.50(c)(1)(ii), states: "Stop use and consult a doctor if the condition persists or gets worse. Do not use longer than 1 week unless directed by a doctor."

In addition, the agency agrees with the comment that alcohol drug products should also bear such a warning. Although a physician may advise using an OTC topical first aid antiseptic for longer than 1 week, consumers should not self-medicate for a longer period of time without consulting a doctor. The Antimicrobial II Panel, in its advance notice of proposed rulemaking for OTC topical antibiotic drug products (42 FR 17642 at 17653), stated that "most small superficial skin wounds including burns, cuts, and abrasions will heal almost completely within 1 week." That Panel expressed concern that "continued use of a topical antibiotic preparation on an unhealed lesion may delay diagnosis and treatment of a more serious skin disease, e.g., a spreading deep bacterial infection, or a wound contaminated with foreign debris such as glass." (42 FR 17653). Because the situation involving use of first aid antiseptics is the same, the warning proposed in this document specifies 1 week rather than 10 days. A 1-week use limit also is consistent with the agency's warning in the final monograph for OTC topical first aid antibiotic drug products. (See 21 CFR 333.150(c)(2).)

25. One comment objected to the number of warnings required for skin wound protectants in proposed § 333.93 (c)(1) through (c)(7). The comment stated that multiple warnings will discourage self-treatment, confuse consumers, and force them to request professional assistance for minor ailments from an overburdened health care distribution system. The comment added that compliance with such lengthy labeling may be difficult because of lack of label space and suggested that, of the seven warnings, only the following two are essential: "For external use only" and "Do not use this product for more than 10 days. If the conditions worsen or persist, see your physician." The comment recommended deletion of all the other proposed warnings because "it is of no benefit to require the appearance of all possible warnings on the label of an over-the-counter medication."

The agency agrees that some of the warnings could be combined or revised without losing their intent. However, limiting the warnings to only the two suggested by the comment would not provide consumers with adequate information. The agency recognizes that it is not necessary or even possible to identify every improper use of a drug that could occur and to list such information on the drug label. Only those warnings that are necessary for the safe and effective use of the product should be included.

The indication for use in this amended tentative final monograph, "First aid to help prevent infection in minor cuts, scrapes, and burns," and the 1-week use limitation warning (see comment 24) should be sufficient to inform the consumer that first aid antiseptics are not to be used on longstanding skin conditions. Therefore, the warning previously proposed in § 333.93(c)(7), "Do not use on chronic skin conditions such as leg ulcers, diaper rash, or hand eczema," is not being included in this amended tentative final monograph.

In addition, the agency has combined and revised the proposed warnings in § 333.93 (c)(3) and (c)(5). (See comment 23.) The proposed warnings in § 333.93 (c)(2) and (c)(4) have also been combined. (See comment 24.) The proposed warning in § 333.93(c)(1) has been retained as suggested by the comment. The proposed warning in § 333.93(c)(6), "Do not use in the eyes," has been expanded to include "or apply over large areas of the body." This revision is in keeping with the agency-initiated change described in the tentative final monograph for OTC first aid antibiotic drug products (see 47 FR 29986 at 29998) that was finalized in the final monograph for those drug products (see 52 FR 47312 at 47324 and 21 CFR 333.50(c)(1)).

The agency believes that these changes will result in labeling that is clear to consumers and that assures safe and proper use of first aid antiseptic drug products.

26. Several comments objected to the warning for antimicrobial bar soaps in proposed § 333.80(c), "Do not use this product on infants under 6 months of age." Some comments recommended deleting the warning and submitted data to show that antimicrobial soaps containing triclosan and triclocarban are safe for use on infants (Ref. 1). Three comments argued that, contrary to the agency's conclusions at 43 FR 1213 and 1232, the data on the use of triclosan in monkey neonates should be regarded as adequate to show that triclosan is safe for human infant use (Ref. 2). The comments further argued that the warning in proposed § 333.80(c) is misleading and will have an unfavorable commercial impact because it will lead consumers to believe that antimicrobial soaps are harmful to users of all ages and therefore consumers will not purchase them. One comment requested that the warning not be required for soap bars weighing 2.5 ounces or less because of the limited space on the label for printing the warning and because these bars probably would not be used on infants over a long period of time.

The labeling section (333.70) in the advance notice of proposed rulemaking (39 FR 33103 at 33141) and the labeling section (333.80) in the tentative final monograph (43 FR 1210 at 1247) entitled "Antimicrobial soap" were intended to apply to antimicrobial bar soaps customarily used in the home. The Panel and the agency recognized that these products were primarily used to "reduce odor" and as "deodorant soaps." No directions for use were proposed in the tentative final monograph because of the known and customary conditions of use. As stated in comment 10, soaps containing antimicrobial ingredients and making only deodorant claims are considered cosmetics and thus are not being included in this amended tentative final monograph. (The regulations governing cosmetics are located in 21 CFR parts 700 to 740.) If the agency determines that cosmetic soap products containing an antimicrobial ingredient need a warning concerning use on infants under 6 months of age, the agency will propose to amend the cosmetic regulations accordingly.

This amended tentative final monograph does not include any products labeled for total body or chronic use in infants. Therefore, the labeling previously proposed in § 333.80, including the warning in § 333.80(c), is not being included in this tentative final monograph.

References

(1) Comment Nos. C00061, CP0002, SUP015, SUP018, C00099, C00109, C00115, and C00134, Docket No. 75N–0183, Dockets Management Branch.

(2) Unpublished Nonclinical Safety Data on Metabolism of Triclosan by Newborn Rhesus Monkeys, Submitted by Ciba-Geigy Corp., Comment No. C00109, Docket No. 75N–0183, Dockets Management Branch.

(3) Published and Unpublished Nonclinical Safety Data on Metabolism of Triclocarban by Infants, Submitted by Armour-Dial, Inc., Comment No. LET047, Docket No. 75N–0183, Dockets Management Branch.

E. Comments on Alcohols

27. Two comments stated that the statement of identity for alcohol drug products proposed by the Miscellaneous External Panel in § 333.98(a), "alcohol for topical antimicrobial use," would be

confusing to consumers. One comment contended that the word "topical" is not generally understood to mean pertaining to the surface, much less to be understood to relate to skin treatment. The comment added that the word "alcohol" in the statement of identity is superfluous because alcohol is already required under section 502(e) of the act (21 U.S.C. 352(e)) to be listed on the label as the active ingredient. The comment pointed out that "antiseptic for the skin" has been the statement of identity for a particular alcohol product since 1928 and that this statement of identity is meaningful to the layman in accordance with 21 CFR 201.61. The comment stated that alcohol and isopropyl alcohol products fit the definition of a skin antiseptic in § 333.3(f) and requested that the indications and directions for use for skin antiseptics in § 333.90 (b) and (d) be used for such alcohol and isopropyl alcohol products.

The other comment argued that the Panel's recommended statement of identity was unnecessary and should be deleted because other sections of the topical antimicrobial monograph already specify that antimicrobial-containing drug products (which would include alcohol and isopropyl alcohol) are to be labeled as skin wound cleansers, antiseptics, etc.

The agency is proposing to include alcohol and isopropyl alcohol in the list of antiseptic active ingredients in § 333.10 of this amended tentative final monograph with the statement of identity "first aid antiseptic." The Miscellaneous External Panel's definition of alcohols in § 333.3(k) is not being proposed in this amended tentative final monograph. Thus "topical," "skin," and "alcohol" are not needed as part of the statement of identity. (See comments 9 and 13.) However, the agency has no objection to these words appearing elsewhere in the labeling of these products as additional information to the consumer, provided they do not appear in any portion of the labeling required by the monograph and do not detract from such required information. (See comment 15.) The indications and directions for "first aid antiseptics" are discussed in comments 16, 17, 20, and 21.

28. One comment argued that the Panel's recommended monograph for alcohol drug products is in conflict with the regulations of the Treasury Department's Bureau of Alcohol, Tobacco and Firearms (BATF) pertaining to ethyl alcohol in 27 CFR parts 211 and 212. (27 CFR parts 211 and 212 were removed in the **Federal Register** of June 2, 1983 (48 FR 24673). Denatured alcohol is now covered in 27 CFR parts 20 and 21.) Under the BATF regulations, denatured ethyl alcohol products containing 70 percent ethyl alcohol are required to be labeled as "Rubbing Alcohol," but under the recommended monograph the identical product could only be labeled as "Alcohol for topical antimicrobial use." The comment pointed out that the Panel itself recognized the effectiveness of alcohol for rubbing uses as well as the fact that these uses had been addressed by another regulatory agency. The comment stated that inconsistency between two regulatory agencies is not sound government policy, is economically unfeasible for manufacturers, and is confusing to consumers. The comment requested that a product that meets the requirements of 27 CFR parts 211 and 212 as well as the requirements of the monograph be allowed to be labeled as a topical antimicrobial product with rubbing indications.

The agency agrees that alcohol drug products for topical antimicrobial use can be labeled, at the option of the manufacturer, to meet both FDA's and BATF's regulations. The appropriate labeling for such a product would include the brand name of the product, if any, and the words "Rubbing Alcohol," in accordance with 27 CFR 211.188 (currently 27 CFR 20.134(e)). This regulation also provides that the manufacturer may include additional statements in the labeling. Thus, the labeling could also contain the words "first aid antiseptic," in accordance with 21 CFR 201.61(b) and proposed § 333.50(a). (See comment 27 for a discussion of "first aid antiseptic" as the statement of identity for these alcohol products.) With this labeling and the labeling proposed in the other parts of § 333.50, a product would meet the requirements of both regulations and provide fully informative labeling to consumers without burdening manufacturers.

29. Noting statements made by the Miscellaneous External Panel (47 FR 22324 at 22327), one comment stated that it appears logical that both alcohol and isopropyl alcohol products should include in their labeling statements to the effect that they "remove dirt and grime" and "do not stain the skin," and that alcohol products should be labeled to "clean and cool the skin" or work "as astringents, counterirritants, or rubefacients." The comment argued that these statements, based on the Panel's report, acknowledge that alcohol products have both cosmetic and medicinal uses and reflect the fact that products with both uses were submitted to the Panel for review.

The agency agrees with the Panel's statements at 47 FR 22327 that alcohols have a variety of uses such as cleaning and cooling the skin. However, these uses are not considered drug uses and as such are not appropriate for inclusion in an OTC drug monograph.

As discussed in comment 16, OTC drug monographs regulate only labeling related in a significant way to the safe and effective use of covered products by lay persons. The statement "does not stain the skin" is not significantly related to the safe and effective use of the product. It is thus outside the scope of the rulemaking, as are statements such as "cleaning and cooling" or "remove dirt and grime." Such statements will be evaluated by the agency on a product-by-product basis, under the provisions of section 502 of the act (21 U.S.C. 352) relating to labeling that is false or misleading. Moreover, any statement that is outside the scope of the monograph, even though it is truthful and not misleading, may not appear in any portion of the labeling required by the monograph and may not detract from such required information. However, terms outside the scope of the monograph may be included elsewhere in the labeling, provided they are not false or misleading.

Because this document addresses only first aid antiseptics, the therapeutic use of alcohol and isopropyl alcohol "as astringents, counterirritants, or rubefacients" will be considered in other rulemakings for external analgesic drug products and skin protectant drug products in future issues of the **Federal Register**. Alcohol and isopropyl alcohol were classified as Category II by the Miscellaneous External Panel in its statement on OTC astringent drug products, published in the **Federal Register** of September 7, 1982 (47 FR 39412 at 39425 and 39436 at 39444). The agency concurred with this classification in the tentative final monograph for OTC astringent drug products, published in the **Federal Register** on April 13, 1989 (54 FR 13490 at 13496).

30. One comment requested the addition of a fourth indication for alcohol active ingredients in proposed § 333.98(b) to allow use as an antibacterial handwash to avoid cross-contamination from one individual to another. The comment argued that products containing alcohols are often used as handwashes by athletic trainers to help prevent the spread of skin

infections from one individual to another in situations in which soap and water are not available, e.g., on the playing field.

Because the scope of this document is limited to first aid products, the indication requested by the comment will not be discussed here. It will be addressed in a future issue of the **Federal Register** covering antimicrobial drug products that are used as antiseptic handwashes.

31. One comment stated that the Advisory Review Panel on OTC Dentifrice and Dental Care Drug Products (Dental Panel) allowed benzocaine or phenol in 70 percent ethyl alcohol for use on the gums (47 FR 22712 at 22737 and 22740). Therefore, it was inconsistent for the Miscellaneous External Panel to place the statement "For application to mucous membranes" in Category II for alcohol drug products (47 FR 22324 at 22332). The comment pointed out that the Miscellaneous External Panel recommended caution in the use of alcohols on mucous membranes in concentrations recommended for antimicrobial use (60 to 95 percent for ethyl alcohol) (47 FR 22327), but the comment did not believe that this caution necessitated an all-inclusive Category II labeling statement. The comment requested that the phrase "except in products containing specific label directions for such use" be added to make the Category II statement read, " 'For application' to mucous membranes, (except in products containing specific label directions for such use)."

An ingredient or drug product can have multiple uses and thus be reviewed by several different panels. The Dental Panel recommended that ethyl alcohol be permitted as a vehicle in concentrations up to 70 percent in products used on the teeth and gums (47 FR 22737 and 22740), but deferred the review of alcohol as an active antiseptic ingredient in the mouth and throat to the Oral Cavity Panel (47 FR 22715). The Oral Cavity Panel placed alcohol in Category III for antimicrobial use in the mouth, but stated that it was ineffective as an antimicrobial agent at concentrations less than 70 percent and that concentrations higher than 35 percent cause burning of mucous membranes (47 FR 22760 at 22872). The Miscellaneous External Panel evaluated ethyl alcohol for use as a topical antimicrobial agent on the skin. The Panel was concerned that alcohol would be irritating to mucous membranes, recommended caution in this use, and placed the statement "For use on mucous membranes" in Category II.

The indications for alcohol drug products covered by this rulemaking apply only to topical antimicrobial uses on the skin and do not include use on mucous membranes, as in the mouth. The agency will address the use of alcohol as an active ingredient on the mucous membranes of the mouth and throat in the proposed rulemaking for oral health care drug products, to be published in a future issue of the **Federal Register**. In developing its proposals in that document, the agency will consider the recommendations of the three Panels, including appropriate concentrations of alcohol in OTC drug products intended for oral use.

32. Two comments requested that small-volume, single-use products containing alcohol active ingredients be exempted from the warning, "Flammable, keep away from fire or flame." The warning was recommended by the Miscellaneous External Panel in § 333.98(c)(1)(ii) (47 FR 22324 at 22330 and 22333). One of the comments argued that swabs saturated with isopropyl alcohol contain such a minute volume of alcohol, 2.5 to 7 mL in each packet, that the warning about flammability is unnecessary for the protection of consumers and may cause undue alarm. The comment pointed out that the United States Department of Transportation excludes such products from the Hazardous Materials Regulations pertaining to flammable liquids.

This comment also requested that the Miscellaneous External Panel's recommended warning in § 333.98(c)(2) for products containing isopropyl alcohol, "Use only in a well-ventilated area; fumes may be toxic," should not be required for single-use alcohol swab products. The comment stated that this warning was proposed by the Panel based on a case in which a large volume of isopropyl alcohol was used in a poorly ventilated room. The comment argued that a large quantity of swabs saturated with isopropyl alcohol would have to be used for this type of application and this is virtually impossible.

The agency disagrees with the comments that small-volume packages containing alcohol active ingredients should be exempted from the flammability warning. The Department of Transportation finding applies only to the shipping of such products in intact packages, whereas the proposed warning informs consumers of proper use after opening the package. The warning is not intended to alarm consumers, but to caution them against improper use of the ingredients. Even small volumes of alcohol should be kept away from fire or flame. The United States Pharmacopeia (U.S.P.) states that isopropyl rubbing alcohol and rubbing alcohol should be labeled to indicate that they are flammable and are to be stored remote from heat (Ref. 1).

However, the agency agrees with the comment that the warning against use in poorly ventilated areas is not needed on small-volume products containing isopropyl alcohol. In fact, the agency tentatively concludes that such a warning is not needed for any product containing isopropyl alcohol because the labeling in this monograph limits its use, i.e., "do not * * * apply over large areas of the body." The agency has reviewed the adverse reactions upon which the Panel based its warning. The three reported cases of adverse effects (Refs. 2, 3, and 4), apparently due to inhalation of isopropyl alcohol, concerned infants in prolonged contact with isopropyl alcohol. The infants were either wrapped in towels saturated with isopropyl alcohol or the alcohol was applied in tepid sponging. The infants were found unconscious or in a stupor after 4 to 8 hours of contact with isopropyl alcohol. Complete recovery occurred on the day following the incident. These three cases, reported between 1953 and 1969, appear to be isolated, infrequent incidents. A warning similar to the one recommended by the Miscellaneous External Panel most probably would not have prevented the adverse reactions reported.

The agency is not proposing that isopropyl alcohol include a warning for toxic fumes in view of the indications provided for in this document, namely, "First aid to help prevent the risk of skin infection in minor cuts, scrapes, and burns." The agency believes that this indication makes it unlikely that anyone using the product as indicated would be exposed to alcoholic fumes for any extended time. Comments are invited on the need for such a warning, including any reports of adverse reactions due to inhalation that have not yet been brought to the agency's attention.

References

(1) "United States Pharmacopeia XXII—National Formulary XVII," United States Pharmacopeial Convention, Inc., Rockville, MD, p. 731, 1989.

(2) Garrison, R.F., "Acute Poisoning from Use of Isopropyl Alcohol in Tepid Sponging," Journal of the American Medical Association, 152:317–318, 1953.

(3) Senz, E.H., and D.L. Goldfarb, "Coma in a Child Following Use of Isopropyl Alcohol in Sponging," Journal of Pediatrics, 53:322–323, 1958.

(4) McFadden, S.W., and J.E. Haddow, "Coma Produced by Topical Application of Isopropanol," Pediatrics, 43:622–623, 1969.

33. One comment requested that the concentration range for ethyl alcohol in proposed § 333.55 (47 FR 22324 at 22332) be broadened to include 48 percent (by volume) aqueous ethyl alcohol for the indications recommended by the Miscellaneous External Panel in § 333.98(b) (47 FR 22330 and 22332). The comment argued that a skin antiseptic does not have to be microbiocidal against all microorganisms, but only against those known to cause infection in minor cuts, scratches, and abrasions.

The comment submitted data on the microbiocidal activity of 48 percent alcohol by volume in aqueous solution and of a marketed product containing the same concentration of alcohol against a variety of micro-organisms, including *Staphylococcus aureus* (*S. aureus*), *Pseudomonas aeruginosa* (*P. aeruginosa*), *Bacillus subtilis* (*B. subtilus*), Proteus species, and *Candida albicans* (*C. albicans*) (Ref. 1). The comment stated that the minimum inhibitory and minimum biocidal activities of the alcohol solution were effective at fourfold and eightfold dilutions and the minimum inhibitory and minimum biocidal activities of the product were effective at eightfold to sixteenfold dilutions, thus indicating effectiveness even if diluted by body fluids at the wound site.

The comment pointed out that, according to the Miscellaneous External Panel, the potential of alcohol to irritate the skin increases with increasing concentration. The comment concluded that an alcohol product should have a high enough concentration to be effective as a skin antiseptic, yet be mild enough to cause minimal skin irritation.

The agency has reviewed the submitted data, which included studies to measure the minimum in vitro contact time for 48 percent alcohol to kill test micro-organisms. Cultures of test micro-organisms were mixed with the test solution containing 48 percent alcohol and subcultured at the following times: 0 (immediately after mixing), 1, 3, 5, 10, and 15 minutes. The test solution killed many test micro-organisms immediately upon contact and all micro-organisms except *B. subtilis* within 1 minute of contact time. The slight increase in time required for 48 percent alcohol to act was insignificant in terms of effectiveness.

Based on these studies and on the advance notice of proposed rulemaking for alcohol drug products (47 FR 22324), the agency proposes that 48 to 95 percent alcohol be classified as Category I. Any authorized formulation of specially denatured alcohol identified in 27 CFR Part 20 may be used. Although the 48-percent alcohol results in an increased time-to-death compared with 60 percent alcohol, the agency believes that the increase in time-to-death is not significant in products for limited first aid antiseptic use.

The agency recognizes that because of its solvent activity, alcohol is frequently used as a vehicle for first aid antiseptic ingredients as well as many other topical medications. As pointed out by the Miscellaneous External Panel (47 FR 22324 at 22327), alcohol is also capable of altering the stratum corneum (skin surface) and enhancing its permeability, thus facilitating the penetration through the skin of any ingredient that is dissolved in it (Ref. 2). For example, enhanced penetration has been demonstrated for iodine (Ref. 3). It is recognized that a wide range of ethyl alcohol concentrations have antiseptic properties (47 FR 22328). However, based upon submitted data for marketed products, only ethyl alcohol in a concentration range of 48 to 95 percent is considered to be an active concentration range for first aid antiseptic use.

The agency notes that the Miscellaneous External Panel included three indications for ethyl alcohol in § 333.98(b): (1) "For first aid use to decrease germs in minor cuts and scrapes," (2) "To decrease germs on the skin prior to removing a splinter or other foreign object," and (3) "For preparation of the skin prior to an injection." Because the agency is now proposing a new first aid antiseptic category for many ingredients, including alcohol, and a general indication, e.g., "First aid to help prevent infection in minor cuts, scrapes, and burns," the agency has not adopted the Panel's first indication. Describing the intent of a product, i.e., "help prevent infection," is more appropriate in a general indication to be included in the monograph than stating a mode of action, i.e., "decreases germs."

The agency believes that the second indication, "to decrease germs on the skin prior to removing a splinter or other foreign object," is a descriptive statement giving an example of a particular kind of first aid. Such illustrative statements are outside the scope of the monograph. Such statements will be evaluated by the agency on a product-by-product basis, under the provision of section 502 of the act (21 U.S.C. 352) relating to labeling that is false or misleading. Moreover, any statement that is outside the scope of the monograph, even though it is truthful and not misleading, may not appear in any portion of the labeling required by the monograph and may not detract from such required information. However, terms outside the scope of the monograph may be included elsewhere in the labeling, provided they are not false or misleading.

The third indication, "For preparation of the skin prior to an injection," will be discussed in a future Federal Register publication on non-first aid uses of antimicrobial ingredients.

References

(1) Comment No. C00142, Docket No. 75N–0183, Dockets Management Branch.

(2) Scheuplein, R. J., and I. H. Blank, "Mechanism of Percutaneous Absorption. IV. Penetration of Nonelectrolytes (Alcohols) from Aqueous Solutions and from Pure Liquids," Journal of Investigative Dermatology. 60:286–296, 1973.

(3) Reeve, T. S., G. A. E. Coupland, and I. B. Hales, "The Effect on Serum Iodine Levels of Painting Tincture of Iodine on the Skin," Medical Journal of Australia, 1: 891–892, 1973.

F. Comments on Chlorhexidine Gluconate

34. Several comments requested that the agency include chlorhexidine gluconate as a Category I ingredient in any amended tentative final monograph. The comments submitted references and data to establish general recognition of safety and effectiveness (Ref. 1) and stated that chlorhexidine gluconate solution is recognized in the "British Pharmacopeia" (Ref. 2) and is formulated in a wide range of products that have been successfully marketed to a material extent and for a material length of time in other countries. The comments asserted that when formulated in compliance with FDA's current good manufacturing practice regulations (21 CFR Part 211), chlorhexidine products are safe and effective for use as skin wound cleansers, skin wound protectants, patient preoperative skin preparations, skin antiseptics, surgical hand scrubs, and health-care personnel handwashes.

A reply comment argued that chlorhexidine gluconate, currently marketed in the United States under approved NDA's, is not eligible for an OTC drug monograph because the ingredient has not been marketed within this country to a material extent and for a material length of time. The comment added that variations in final formulations may alter the safety and effectiveness of the ingredient. The comment submitted data (Ref. 3) to support this viewpoint and requested that chlorhexidine gluconate be classified in Category II.

In the previous tentative final monograph (43 FR 1210), chlorhexidine gluconate (4 percent solution) was neither addressed nor categorized as Category I, II, or III. However, subsequent to the tentative final monograph, the agency granted a petition (Ref. 4) and in the **Federal Register** of March 9, 1979, reopened the administrative record to allow interested persons an opportunity to submit data and information (44 FR 13041). The comments (Ref. 1) and reply comment (Ref. 2) were submitted in response to that notice. However, since that time a majority of the comments on chlorhexidine submitted in response to the notice have been withdrawn (Ref. 5). While the data and information remain on public display as part of the administrative record, they are no longer being considered in this rulemaking.

The agency has reviewed the marketing history of chlorhexidine gluconate and finds that although it has been marketed for professional or hospital use, this ingredient has never been marketed in the United States for any first aid use. Accordingly, chlorhexidine gluconate 4 percent aqueous solution as a first aid antiseptic is a new drug and is not included in this proposed monograph.

The professional uses for chlorhexidine gluconate requested by the comments (Ref. 1), i.e., surgical hand scrub and health-care personnel handwash, will be addressed separately in the segment of this rulemaking dealing with uses other than first aid in a future issue of the **Federal Register.**

References

(1) Comment Nos. C00110, C00116, C00120, C00130, C00131, C00136, C00137, EXT018, RC0002, RC0005, CP0003, LET012, LET014, LET016, SUP030, SUP033, SUP038, and SUP040, Docket No. 75N–0183, Dockets Management Branch.

(2) "British Pharmacopeia," Vol. I, Her Majesty's Stationery Office, London, pp. 100–101, 1980.

(3) Comments No. RC001 and RC004, Docket No. 75N–0183, Dockets Management Branch.

(4) Citizen Petition No. CP003, Docket No. 75N–0183, Dockets Management Branch.

(5) Comments No. WDL003, WDL004, and WDL005, Docket No. 75N–0183, Dockets Management Branch.

G. Comments on Chloroxylenol

35. A number of comments disagreed with the agency's Category III classification of chloroxylenol in the tentative final monograph. They argued that reevaluation of the data previously submitted to the agency along with new data that have been submitted (Refs. 1 through 16) would provide adequate justification for classifying chloroxylenol in Category I for safety and effectiveness for use in antimicrobial soaps, health-care personnel handwashes, patient preoperative skin preparations, skin antiseptics, skin wound cleansers, skin wound protectants, and surgical hand scrubs. Several comments pointed out that the Antimicrobial II Panel unanimously concluded that chloroxylenol is generally recognized as safe for topical use in athlete's foot and jock-itch preparations. One comment stated that the Panel placed hexylresorcinol in Category I and chloroxylenol in Category III as a skin wound cleanser, but that a comparison of the available data clearly indicates that the safety data available on chloroxylenol are superior to those for hexylresorcinol.

Data submitted by the comments regarding safety and effectiveness for uses other than first aid, e.g., health-care personnel handwash, patient preoperative skin preparation, and surgical hand scrub will be discussed in the segment of this rulemaking dealing with uses other than first aid in a future issue of the **Federal Register.**

In the tentative final monograph, chloroxylenol was categorized as Category III for safety and effectiveness as a skin antiseptic, skin wound cleanser, and skin wound protectant, and it was recommended that effectiveness testing, both in vitro and in vivo, be done (43 1210 at FR 1238). The agency also requested data to show the effects of chloroxylenol on wound healing (43 FR 1238).

Subsequent to the tentative final monograph, the Antimicrobial II Panel in the advance notice of proposed rulemaking for OTC topical antifungal drug products categorized chloroxylenol (0.5 to 3.75 percent) as safe (Category I) for short-term use (up to 13 weeks) (47 FR 2480 at 12535).

The agency has reviewed the data, which include wound-healing studies, submitted by the comments. Bradbury and Hayden (Refs. 9 and 10) described studies on the effect of various concentrations of chloroxylenol up to 4.8 percent, on wound healing in rats. Wound healing was assessed by measuring wound tensile strength and histopathology. The results showed that none of the treatments significantly altered wound tensile strength or caused a significant delay in the healing process.

Maibach (Refs. 11 and 22) described two clinical studies that used the forearms of human volunteers to assess the effects of petroleum jelly and carbolated petroleum jelly, containing chloroxylenol 0.5 percent, on wound healing. In one study (Ref. 11), the forearm skin was stripped and treated twice daily for 5 days. In the other study, incisions were made in the forearm and treated three times in 24 hours (Ref. 12). There were no differences in the rates of wound healing between control sites and treated sites.

These studies (Refs. 9 through 12) showed that chloroxylenol 0.5 percent to 4.8 percent did not delay wound healing and affirm the Antimicrobial II Panel's conclusion that chloroxylenol is safe for short-term use. Accordingly, the agency is reclassifying chloroxylenol to Category I for safety for use as a first aid antiseptic.

The in vitro data demonstrate that formulated chloroxylenol, in the presence of 5 percent serum (37 °C) is effective within 5 to 10 minutes. The in vivo data, derived from studies of artificial contaminants on the skin of human test subjects, showed that chloroxylenol-containing product reduced the number of staphylocci, pseudomonas, escherichia, and streptococci by greater than one log (i.e., 1 $\log_{10}$) within 5 minutes. However, none of the studies demonstrate the contribution of chloroxylenol to the formulated product.

The agency does not consider the data regarding the antiseptic activity of chloroxylenol itself to be adequate. While the data are considered sufficient to support in vitro and in vivo effectiveness for the finished products (Refs. 13 through 16), the available data are inadequate to show the contribution of the chloroxylenol. Because these finished products contain several additional ingredients, i.e., surfactants, isopropanol, pine oil, or ethylenediaminete-tracetic acid (EDTA), any of which could have contributed germicidal activity, conclusions regarding chloroxylenol's active contribution to the products' efficacy cannot be supported. Accordingly, in this proposed rule chloroxylenol is being proposed as a Category III first aid antiseptic ingredient for effectiveness.

References

(1) Unpublished Clinical Safety and Effectiveness Studies on Aqueous Soap Formulations, submitted by the Pennwalt Corp., Comment No. 0B0007, Docket No. 75N–0183, Dockets Management Branch.

(a) Controlled Clinical Study Comparing the Activity of Fresh, Camay Soap, and Phisohex Against the Natural Bacterial Flora of the Hand.

(b) Antimicrobial Activity of PCMX, Triclosan, and TCC.

(c) Repeated Insult Patch Testing of Fresh Soap.

(2) Unpublished Nonclinical and Clinical Studies, and Protocols, submitted by the

Pennwalt Corp., Comment No. C00096, Docket No. 75N–0183, Dockets Management Branch.

(a) Part I: PCMX Toxicosis, final reports of completed studies, interim reports of incomplete studies, and Preclinical Testing Protocol.

(b) Part II: Complete Reports on Clinical Safety and Efficacy and In Vitro Efficacy Studies.

(3) Unpublished Clinical Effectiveness Studies on Aqueous Soap Formulations, submitted by Chemical Specialties, Inc., Comment No. C00122, Docket No. 75N–0183, Dockets Management Branch.

(a) Protocol and Results of a Glove Juice Hand Washing Test Performed with PHLO Antimicrobial Skin Cleanser.

(b) Results of a Zone of Inhibition and Assay Performed on Aged Samples of PHLO Antimicrobial Skin Cleanser.

(4) Unpublished Clinical Safety and Effectiveness Studies on Aqueous Soap Formulations, submitted by Sani-Fresh, Comment No. C00123, Docket No. 75N–0183, Dockets Management Branch.

(a) Bactericidal Activity of Envair Antiseptic Hand Soap.

(b) Dermal Irritation Study.

(c) Insult Patch Test.

(d) Bacterial Kill Test.

(e) Hand-wash Effectiveness Test.

(5) Unpublished In Vitro Effectiveness Studies Performed on Aqueous Soap Solutions, submitted by Seagull Chemical, Comment No. C00125, Docket No. 75N–0183, Dockets Management Branch.

(a) AOAC Available Chlorine Germicidal Equivalent Concentration Test.

(b) The Antimicrobial Activity of a Sample.

(6) Published and Unpublished Nonclinical and Clinical Safety Studies, submitted by Ferro Corp., Comment No. SUP011, Docket No. 75N–0133, Dockets Management Branch.

(7) Published and Unpublished Safety and Effectiveness Studies, submitted by Scientific and Regulatory Services, Comment No. SUP012, Docket No. 75N–0183, Dockets Management Branch.

(8) Unpublished Clinical Safety and Effectiveness Studies, submitted by Chesebrough Ponds, Inc., Comment No. SUP010, Docket No. 75N–0183, Dockets Management Branch.

(a) The Effects of Vaseline Petroleum Jelly and Vaseline First Aid Carbolated Petroleum Jelly on Epidermal Wound Healing—A Controlled Clinical Laboratory Study, April 29, 1976.

(b) The Effect of Vaseline Petroleum Jelly and Vaseline First Aid Carbolated Petroleum Jelly on Healing of Experimental Skin Wounds, January 13, 1977.

(9) Bradbury, S.J., and J. Hayden, "Effect of Dettol® on Wound Healing in Rats," Report No. RC 76132, unpublished study, Comment No. SUP05, Docket No. 75N–0183, Dockets Management Branch.

(10) Bradbury, S.J., and E.J. Hayden, "Dettol® Wound Healing," unpublished study, Project No. RC 1081, 1978, Comment No. SUP012, Docket No. 75N–0183, Dockets Management Branch.

(11) Maibach, H.I., "The Effects of Vaseline® Petroleum Jelly and Vaselines First Aid Carbolated Petroleum Jelly on Epidermal Wound Healing—A Controlled Clinical Laboratory Study," unpublished study, Comment No. SUP010, Docket No. 75N–0183, Dockets Management Branch.

(12) Maibach, H.I., "The Effect of Vaseline® Petroleum Jelly and Vaseline®First Aid Carbolated Petroleum Jelly on Healing of Experimental Skin Wounds," unpublished study, Comment No. SUP010, Docket No. 75N–0183, Dockets Management Branch.

(13) Munton, T.J., and J. Prince, "The Bacteriostatic and Bactericidal Activity of Dettol® Against a Range of Recently Isolated Mesophilic Strains Including Members of the Normal Flora and Cutaneous Pathogens of the Skin," unpublished study, No. BL 75/4, 1975, Comment No. SUP003, Docket No. 75N–0183, Dockets Management Branch.

(14) Prince, J., and K.A. Barker, "A Comparison of the In-Vitro Activity of Dettol®, Hexylresorcinol and Benzalkonium Chloride," unpublished study, No. BL 76/28, 1976, Comment No. SUP003, Docket No. 75N–0183, Dockets Management Branch.

(15) Munton, T.J., and J. Prince, "The Bactericidal Activity of Dettol® on Skin Artificially Contaminated with Micro-organisms Using the Replica Plating Technique," unpublished study, No. BL 75/14, RC 7565, 1975, Comment No. SUP003, Docket No. 75N–0183, Dockets Management Branch.

(16) "Scientific Information on the 'In-vitro' and 'In-vivo' Antimicrobial Activity of Dettol® as Determined in the Bacteriological Laboratories of Reckitt and Colman, Hull," published report, Comment No. C00062, Docket No. 75N–0183, Dockets Management Branch.

H. Comments on Hydrogen Peroxide

36. Two comments requested that hydrogen peroxide solution (3 percent) be included in the monograph as a Category I skin antiseptic. One comment pointed out that no mention of the ingredient is made in the proposed or tentative final monograph even though hydrogen peroxide has been recognized by the U.S.P. for many decades as a topical anti-infective for application to skin and mucous membranes. The comment submitted two references to show that hydrogen peroxide is a desirable skin antiseptic that can be used safely and effectively by the layman (Refs. 1 and 2).

Hydrogen peroxide solution (3 percent) for use as a skin antiseptic was not classified in the previous tentative final monograph because it was deferred to the Miscellaneous External Panel. (See comment 85, 43 FR 1210 at 1223.) A manufacturer had made a submission (Ref. 3) on hydrogen peroxide (U.S.P., 3 percent) as a first aid antiseptic drug product to the Miscellaneous External Panel but that Panel disbanded before it reviewed hydrogen peroxide. The agency subsequently concluded that it would be appropriate to categorize hydrogen peroxide as a first aid antiseptic in this antimicrobial rulemaking. Accordingly, the agency requested and received permission from the manufacturer to place the manufacturer's submission (Ref. 3) on public display in the Dockets Management Branch under the antimicrobial docket number (Ref. 4).

The submission forwarded by the manufacturer (Ref. 3) included labeling for a currently marketed product containing hydrogen peroxide solution U.S.P. 3 percent, which states: "First aid antiseptic" "For treatment of minor cuts and abrasions." The submission also included safety and effectiveness data from published articles and unpublished studies. These data indicate that hydrogen peroxide inhibits *S. aureus, Salmonella typhosa, Escherichia coli (E. coli), Proteus vulgaris, Klebsiella pneumoniae, Streptococcus hemolyticus,* and *P. aeruginosa.* The manufacturer also provided in vitro data to show that 3 percent hydrogen peroxide reduced the number of *S. aureus* ATCC 6538P by 3 logs (3 $\log_{10}$) within 5 minutes and completely inhibited all bacteria within 10 minutes.

In a separate OTC drug rulemaking, for OTC oral mucosal injury drug products, the agency found hydrogen peroxide (3 percent in aqueous solution) safe for short-term use up to 7 days. (See the **Federal Register** of July 26, 1983, 48 FR 33984 at 33993.)

Hydrogen peroxide achieves its intended benefit in vivo by means of both a mechanical action and a measurable antibacterial action. Because hydrogen peroxide has been demonstrated to be both safe and effective for use in minor wounds, the agency is proposing to classify hydrogen peroxide (3 percent in aqueous solution) as Category I for use as a first aid antiseptic drug product.

References

(1) "Antiseptics and Disinfectants," *in* "AMA Drug Evaluations," 2d Ed., American Medical Association, Publishing Sciences Groups, Inc., Ashton, MA, p. 653, 1973.

(2) Schumb, W. C., C. N. Satterfield, and R. L. Wentworth, "Hydrogen Peroxide," American Chemical Society Monograph Series, 128, Reinhold Publishing Corp., New York, 1955.

(3) OTC Volume 160031.

(4) Letter from M. Kaplan, Parke-Davis, Division of Warner Lambert and Co., to W. E. Gilbertson, FDA, dated July 12, 1982, Comment No. LET051, Authorizing Public Display of OTC Volume No. 160031, Docket No. 75N–0183, Dockets Management Branch.

I. Comments on Iodine and Iodophors

37. One comment objected to the classification of iodine tincture in Category III for use as a skin antiseptic. To justify Category I status, the comment cited the more than 130-year

history of use of iodine tincture as a household first aid product and the extensive literature on iodine as an antiseptic published during the past several decades. The comment submitted two studies to support its position (Refs. 1 and 2). According to the comment, the study by Salle and Catlin (Ref. 1) showed that iodine tincture (2 percent) has the highest germicidal activity and the lowest toxicity of the germicides tested. The comment pointed out that the publication by Gershenfeld and Witlin (Ref. 2) concluded that iodine was a highly effective bactericidal agent against many different species of microorganisms at high dilution and within a wide pH range; and that it possessed a very low toxicity to tissues as determined by many varied in vitro and in vivo toxicity tests, including tests on human skin. The comment added that an extensive list of additional references has been included as part of the cited studies, and that these references should help resolve the questions raised by the Commissioner. The comment recommended that iodine tincture be placed in Category I.

In the tentative final monograph, the agency concluded that elemental iodine hydroalcoholic solution (iodine tincture) is effective for first aid use on minor wounds as a skin antiseptic, skin wound protectant, and skin wound cleanser, although questions remained regarding the minimally effective dose and the effect of organic load and pH. In addition, the agency was concerned about the irritating properties of iodine and delay in wound healing and therefore classified iodine tincture in Category III (43 FR 1210 at 1234).

The agency has reviewed the data and information submitted by the comments (Refs. 1 and 2), which described reports from studies on the properties of elemental iodine, iodine tincture U.S.P., and iodine solution U.S.P. The studies, not previously reviewed in either the Panel report or in the tentative final monograph, provided data primarily pertaining to effectiveness.

The agency has also considered additional studies in test wounds of laboratory animals. Branemark et al. (Refs. 1 and 2), which described reports from studies on the properties of elemental iodine, iodine tincture U.S.P., and iodine solution U.S.P. The studies, not previously reviewed in either the Panel report or in the tentative final monograph, provided data primarily pertaining to effectiveness.

The agency has also considered additional studies in test wounds of laboratory animals. Branemark et al. (Ref. 3) inflicted minute test wounds and control wounds in the skin of mice, hamsters, and rabbits. The test wounds were treated with iodine solutions, and the structure of the skin was observed microscopically for healing. Various antiseptic ingredients, including iodine in saline solution, were tested on minute cutaneous wounds. Microscopic analysis showed very slight tissue injury from the antiseptic.

Edlich et al. (Ref. 4) inflicted deep wounds in the skin of guinea pigs, contaminated the wounds with *S. aureus*, waited 5 minutes, cleansed the wounds with 100 milliliters (mL) of various antiseptic solutions, including 70 percent alcohol, iodine aqueous solution or iodine tincture, and saline control solutions. The wounds were closed with tape, observed, and measured for inflammatory responses (i.e., induration and pus). Subcultures were made for viable bacteria. Edlich et al. reported that 70 percent alcohol, iodine aqueous solution, and iodine tincture helped to reduce the rate of infection without causing significant inflammatory responses in the wounds. Specifically, the authors stated that "The gross infection score, the indurated wound margin, and the percentage of positive cultures in the contaminated wound receiving a single irrigation with tincture of iodine were significantly less than the corresponding inflammatory responses in the control wounds."

Based on the available data, the agency concludes that 2 percent aqueous or alcoholic solutions of elemental iodine (i.e., iodine tincture, U.S.P. or iodine topical solution, U.S.P.) are safe and effective for first aid use to decrease the number of bacteria in minor cuts and scrapes. Therefore, the agency is proposing that these iodine solutions be Category I for use as a first aid antiseptic.

References

(1) Salle, A. J., and B. W. Catlin, "Profile Evaluations of Germicides," *Journal of the American Pharmaceutical Association*, (Scientific Edition), 36:129–133, 1947.

(2) Gershenfeld, L., and B. Witlin, "Iodine as an Antiseptic," Annals of the New York Academy of Sciences, 53:172–182, 1950.

(3) Branemark, P. J., et al. "Tissue Injury Caused by Wound Disinfectants," Journal of Bone and Joint Surgery, 49:48–62, 1967.

(4) Edlich, R. F., et al., "Studies in Management of the Contaminated Wound. III. Assessment of the Effectiveness of Irrigation with Antiseptic Agents," The American Journal of Surgery, 118:21–30, 1969.

38. A number of comments submitted new data (Ref. 1) to establish that povidone-iodine is safe and effective as a topical antimicrobial drug. The comments requested that povidone-iodine be reclassified from Category III to Category I as a topical antimicrobial ingredient for use as an antimicrobial soap, health-care personnel handwash, surgical hand scrub, patient preoperative skin preparation, skin antiseptic, skin wound cleanser, and skin wound protectant.

The agency has considered the new povidone-iodine data submitted in support of the request to reclassify povidone-iodine from Category III to Category I as well as the reports of other advisory panels. On the basis of this information FDA has tentatively concluded that povidone-iodine should be classified in Category I for use as a first aid antiseptic.

The Advisory Review Panel on OTC Contraceptives and Other Vaginal Drug Products, in its report published October 13, 1983, stated that "microbiocidal effectiveness of povidone-iodine has been clearly demonstrated by in vitro studies against a variety of pathologic bacteria, fungi, and protozoan organisms" and "in clinical studies, povidone-iodine has been shown to disinfect skin and mucous membrane" (48 FR 46694 at 46705). That Panel classified povidone-iodine diluted to 0.15 to 0.30 percent for use as a douche as Category I for the "relief of minor irritation of the vagina" but reserved directions for use of full-strength solution for professional uses.

The Antimicrobial II Panel reviewed povidone-iodine as a topical antifungal ingredient. In its evaluation, the Panel relied on new safety data as well as the recommendations of the Antimicrobial I Panel in the **Federal Register** published September 13, 1974 (39 FR 33103 at 33129). The Antimicrobial II Panel's recommendations on antifungal use of povidone-iodine were published in the March 23, 1982 **Federal Register** (47 FR 12480 at 12545) as an advance notice of proposed rulemaking. That Panel concluded that 10 percent povidone-iodine was safe for OTC topical antifungal use in the treatment of athlete's foot, jock itch, and ringworm.

The safety aspects of povidone-iodine as a topical first aid antiseptic for consumer use in the home environment (short-term use over limited areas of the skin) are essentially the same as those described by the Antimicrobial II Panel for topical antifungal ingredients. The agency concurs with and adopts the Antimicrobial II Panel's safety evaluation of povidone-iodine. Povidone-iodine is being proposed as generally recognized as safe as an OTC topical first aid antiseptic ingredient in this amended tentative final monograph. (See comments 41 and 42 for additional safety discussions. See comment 39 for effectiveness discussion.)

Reference

(1) Comments No. C00104, C00108, C00111, C00112, C00113, C00128, C00132, and C00133, Docket No. 75N–0183, Dockets Management Branch.

39. Several comments requested that the tentative final monograph specify the lowest potency concentration of available iodine that marketed preparations be allowed to reach before

being considered ineffective and, thus, adulterated or misbranded (Refs. 1 to 4). One comment (Ref. 2) asserted that "many noncompendial povidone-iodine preparations do not specify the labeled amount of iodine, and there is wide variation in their potency. This has created confusion in the market and may put consumers at risk." The comment requested "that those preparations which are placed in Category I contain in the respective use monograph a lower potency limit, irrespective of the original concentration, since this lower limit would still be effective." Another comment (Ref. 3) suggested that the monograph be revised to include povidone-iodine as an antimicrobial bar soap containing not less than 5 percent nor more than 10 percent povidone-iodine U.S.P., equivalent to 0.5 percent and 1.0 percent available iodine. Topical dosage for use as a solution containing not less than 7.5 percent nor more than 10 percent povidone-iodine U.S.P., equivalent to 0.75 percent and 1.0 percent available iodine, was proposed for a health-care personnel hand wash, surgical hand scrub, skin antiseptic, skin wound cleanser, skin wound protectant, or a patient preoperative skin preparation.

One comment (Ref. 4) included data on the rate of release of iodine from povidone-iodine to support effectiveness.

In the previous tentative final monograph, the agency did not discuss or recommend specific concentrations of povidone-iodine for the proposed seven classes of preparations (i.e., antimicrobial soap, health-care personnel handwash, surgical hand scrub, skin antiseptic, skin wound cleanser, skin wound protectant, and patient preoperative skin preparation) (43 FR 1210 at 1235). However, the agency stated that "the question of iodine release from the complexed molecule, including rate of release and binding to other materials, as well as the influence of the release rate on effectiveness, must be resolved" (43 FR 1236).

Subsequently, the agency has reviewed chemical data and in vivo and in vitro biological data that support the effectiveness of povidone-iodine (Refs. 1 through 4). The biological data show that dilutions from marketed 5 percent povidone-iodine and marketed 7.5 to 10 percent povidone-iodine significantly reduced the number of test bacteria within 1 minute (Refs. 1 and 2). According to references that were submitted in connection with another rulemaking, povidone-iodine solution at concentrations of 1 to 10 percent contains over 99 percent complexed iodine (Ref. 5). Based on an iodine-starch reaction as a biological model, it has been shown that any iodine that is removed from the complex would be replaced within less than 25 milliseconds (Ref. 6). The agency's detailed evaluation is on display in the Dockets Management Branch (Ref. 7).

The data show that as the already released iodine interacts chemically with the microbes, more iodine is rapidly released from the povidone-iodine. Consequently, the availability of the iodine is not a problem. Furthermore, povidone-iodine manufactured in accordance with current good manufacturing practices (21 CFR Part 211) should not present problems. Based on the available data, povidone-iodine at 5 to 10 percent concentrations is being classified as Category I for first aid antiseptic use.

Other uses for povidone-iodine will be addressed separately in the segment of this rulemaking dealing with uses other than first aid in a future issue of the **Federal Register**.

References

(1) Comment No. C00104, Docket No. 75N–0183, Dockets Management Branch.

(2) Comment No. C00128, Docket No. 75N–0183, Dockets Management Branch.

(3) Comment No. C00103, Docket No. 75N–0183, Dockets Management Branch.

(4) Comment No. C00111, Docket No. 75N–0183, Dockets Management Branch.

(5) Schenck, H.U., et al., "Structure of Povidone-Iodine," *in* "Current Chemotherapy and Infectious Disease," Volume I, American Society for Microbiology, Washington, pp. 477–478, 1980.

(6) Ditter, W., D. Horn, and E. Luedekke, "Thermodynamic and Kinetic Examinations Concerning the Complex Binding State and the Rate of Liberation of Iodine from Aqueous Iodine-PVP-Solutions," included in Comment No. C00012, Docket No. 81N–0014, Dockets Management Branch.

(7) Letter from W. E. Gilbertson, FDA, to L. Blecher, GAF Corp., October 5, 1983, Coded LET004, Docket No. 81N–0114, Dockets Management Branch.

40. Several comments objected to FDA's requiring expiration dates (not to exceed 2 years after manufacture) for all products containing an iodophor active ingredient (43 FR 1210 at 1235). The comments stated that stability varies among different iodophor products, with some products falling short of, and others far exceeding, this time period. The comments argued that data derived from a particular formulation are not applicable to other iodophor categories or even to formulations containing a common active ingredient because of the nature of the particular formulation, the purity of the active ingredient, other substances used, and the level of manufacturing expertise available.

The comments also pointed out that the fixed 2-year time period is contrary to FDA's policy under the good manufacturing practice regulations, which require that the stability profile of each individual product in its own container-closure system under varying environmental conditions be known and controlled. The comments argued that it is important that expiration dating for iodophor products be supported by each manufacturer, with well-defined test data, for the stability term that is proposed for a particular formulation and that such support data should include studies conducted under conditions of actual use demonstrating that the formulation is stable for the period claimed.

The agency agrees with the comments. At the time the agency proposed expiration dates, the agency was concerned with the lack of stability data submitted for the several iodophor preparations. However, current good manufacturing practice regulations (21 CFR parts 210 and 211) require a testing program to assess the stability of finished products and to determine appropriate storage conditions and an expiration date. Under § 211.137, drug products must bear an expiration date supported by reliable stability data. The agency has proposed an exemption from this requirement for human OTC products that do not bear dosage limitations if appropriate data show that the products are stable for at least 3 years. (See § 211.137(g).) Because of these current good manufacturing practice regulations, the agency concludes it is not necessary to include specific expiration dating periods for dosage forms of povidone-iodine or other iodophors in the amended tentative final monograph. Therefore, the previously proposed 2-year expiration date for iodophors has been eliminated.

41. Several comments submitted data from published and unpublished studies to show that povidone-iodine does not alter thyroid function (Refs. 1 through 12). These data were submitted in response to FDA's request for controlled research to show the conditions of use under which thyroid function would, or would not, be altered (43 FR 1210 at 1235). The comments stated that although the data show that the amount of total serum iodine, or iodide, is increased after povidone-iodine is used topically, there is no significant alteration of the level of thyroid hormone measured by RT_3U, T_3, and T_4 radioimmunoassays.

One comment pointed out that, as part of the ongoing review of food additives, FDA issued a final order on March 31, 1976 (43 FR 11699) (21 CFR 184.1634) confirming that potassium iodide, a salt of iodine, is generally recognized as safe. The comment also pointed out that, in a separate rulemaking procedure for OTC vitamin and mineral drug products (44 FR 16126 at 16181), the Advisory Review Panel on OTC Vitamin, Mineral, and Hematinic Drug Products discussed the safety of iodine and advised that the thyroid can safely absorb up to 2 milligrams of iodine without metabolizing it. This absorption prevents the accumulation of iodine that would inhibit thyroid hormone synthesis. The comment added that the administrative record for the antimicrobial monograph contains adequate data to show that topically applied iodine is virtually not absorbed.

The agency has reviewed the data submitted (Refs. 1 through 12) and agrees with the comments that thyroid dysfunction does not occur from the topical use of povidone-iodine. Plasma iodine levels may be elevated following the topical use of povidone-iodine; however, the thyroid adapts to the iodine elevation, and the iodine is readily excreted by the kidney without evidence of thyroid dysfunction. During a study of the effects of surgical scrubbing with povidone-iodine for 2 weeks, it was concluded that some absorption of iodine did occur when povidone-iodine was used topically (Ref. 12). The serum iodine concentration was elevated, but not protein-bound iodine, T_4, T_3, or TSH. However, the level of serum iodine returned to normal when povidone-iodine use was discontinued. In addition, studies following the application of povidone-iodine to the mucous membranes (vagina) and intact and damaged skin in humans and animals reported protein-bound iodine elevations, but no alterations in thyroid function (Refs. 4, 7, 9, and 10). Therefore, the agency believes that topically applied povidone-iodine does not cause thyroid dysfunction and is safe for OTC use.

References

(1) Alden, E.R., et al., "Effect of Prenatal Povidone-Iodine Perineal Antisepsis on Serum Protein-Bound Iodine," Obstetrics and Gynecology, 35:253–254, 1970.

(2) Garnes, A.L., et al., "Clinical Evaluation of Povidone-Iodine Aerosol Spray in Surgical Practice," American Journal of Surgery, 97:49–53, 1959.

(3) Goldman, M., and D. Landry, "The Effect of Povidone-Iodine on Thyroid Function in Rats," Toxicology and Applied Pharmacology, 35:341–346, 1976.

(4) Gortz, G., "Povidone-Iodine (Mundidone)—Alternative to Topical Antibiotics: Effects and Side Effects in Wound Treatment," 10th International Congress of Chemotherapy, Zurich, 1977.

(5) Higgins, H.P., et al., "The Effect of Povidone-Iodine (Betadine) on Serum Protein-Bound Iodine, When Used as a Surgical Preparation on Intact Skin," The Canadian Medical Association Journal, 90:1298–1300, 1964.

(6) Kearns, J.E., "The Effect of New Iodophors on Protein-Bound Iodine and Butinol Extractable Iodine in Humans," American Journal of Surgery, 109:457–459, 1965.

(7) King, I.R., and A.W. Diddle, "Protein-Bound Iodine and T_4 Tests After Vaginal Application of Povidone-Iodine," American Journal of Obstetrics and Gynecology, 108:1175–1177, 1970.

(8) Lavelle, K.J., et al., "Iodine Absorption in Burn Patients Treated Topically with Povidone-Iodine," Clinical Pharmacology and Therapeutics, 17:355–362, 1975.

(9) Meissner, K., et al., "Povidone-Iodine Versus Antibiotic Application in Prophylaxis and Treatment of Peritonitis: Effects on Thyroid Function," 10th International Congress of Chemotherapy, Zurich, 1977.

(10) Quagliana, J.M., "Effect of Topical Povidone-Iodine (Betadine) on Serum Protein-Bound Iodine," Journal of Clinical Endocrinology and Metabolism, 23:395–397, 1963.

(11) Renk, E., et al., "The Influence of Povidone-Iodine (Mundidone) on the PBI and BEI Serum Levels in Burn and Peritonitis Therapy," 10th International Congress of Chemotherapy, Zurich, 1977.

(12) Ingbar, S.H., "Studies of the Effects of Surgical Scrubbing with PVP–I," unpublished study included in Comment No. C0032, Docket No. 75N–0183, Dockets Management Branch.

42. Several comments objected to the Commissioner's statement in the antimicrobial tentative final monograph that data presented to the Panel suggested that nonsurfactant iodophor products (povidone-iodine) delay the rate of wound healing (43 FR 1210 at 1235). One comment submitted new data to show that povidoneiodine has no adverse effect on wound healing in animals or humans (Refs. 1 through 13). Another comment stated that povidoneiodine may, in fact, aid wound healing.

The agency has reviewed the new data submitted by the comments and agrees that povidone-iodine does not delay wound healing. Controlled studies on wound healing were conducted in animals and humans and involved various types of dermal wounds and several antiseptics, including povidone-iodine. Both superficial and deeper wounds were studied with a contralateral control, and clinical evaluation was also done on patients receiving split-skin grafts. Results showed that there were no statistically significant differences in mean healing times between any of the treatment groups and their saline controls. In addition, microscopic analysis showed no differences in wound healing in the groups studied. These pathological and histological studies did not indicate any deleterious effect of povidone-iodine on wound healing. However, there was also no evidence demonstrating that povidone-iodine might aid wound healing.

References

(1) Paster, Z., "A Study of the Effect of Polydine on Wound Healing," Israel Institute for Biological Research, 1977, unpublished study, EXT012, Docket No. 75N–0183, Dockets Management Branch.

(2) Gruber, R.P., L. Vistnes, and R. Pardoe, "The Effect of Commonly Used Antiseptics on Wound Healing," Plastic and Reconstructive Surgery, 55:472–476, 1975.

(3) Gilmore, O.J.A., "A Reappraisal of the Use of Antiseptics in Surgical Practice," Annals of the Royal College of Surgeons of England, 59:93–103, 1977.

(4) Gilmore, O.J.A., C. Reid, and A. Strokon, "A Study of the Effect of Povidone-Iodine on Wound Healing," Postgraduate Medical Journal, 53:122–125, 1977.

(5) Sindelar, W.F., and G.R. Mason, "Irrigation of Subcutaneous Tissue with Povidone-Iodine Solution for Prevention of Surgical Wound Infections," Surgery Gynecology and Obstetrics, 148:227–231, 1979.

(6) Gilmore, O.J.A., "Prevention of Wound Infection," Lancet, 1:1134, 1973.

(7) Gilmore, O.J.A., and T.D.M. Martin, "Aetiology and Prevention of Wound Infection in Appendectomy," British Journal of Surgery, 61:281–287, 1974.

(8) Gilmore, O.J.A., et al., "Colonic Anastomosis Healing: The Effect of Topical Povidone-Iodine," European Surgical Research, 10:94–104, 1978.

(9) Gilmore, O.J.A., and P.J. Sanderson, "Prophylactic Interparietal Povidone-Iodine in Abdominal Surgery," British Journal of Surgery, 62:792–799, 1975.

(10) Gilmore, O.J.A., "Intraperitoneal Povidone-Iodine," Lancet, 2:37–38, 1977.

(11) Morgan, W.J., "Povidone-Iodine Spray for Wounds Sutured in the Accident Department," Lancet, 1:769, 1978.

(12) Gilmore, O. J. A., et al., "Prophylactic Intraperitoneal Povidone-Iodine in Alimentary Tract Surgery," American Journal of Surgery, 35:156–159, 1978.

(13) Gilmore, O. J. A., "Experimental Treatment of Peritonitis and Peritoneal Adhesions with Antiseptics," in "Proceedings of the World Congress on Antiseptics," Limburg/Lahm, Germany, pp. 117–119, 1976.

43. Several comments requested clarification of contradictory statements concerning the compatibility of iodophors in antimicrobial soaps. The agency agreed to delete the statement of incompatibility of povidone-iodine in soap formulation (43 FR 1210 at 1221; comment 70), but then at 43 FR 1236 the agency stated that it was unaware of any data to show that iodophors can be

formulated into antimicrobial soaps. Another comment pointed out that the agency's conclusion at 43 FR 1236 was inconsistent with the list of Category III active ingredients at 43 FR 1229. One comment also requested that poloxamer-iodine complex be deleted from the Category II list for antimicrobial soaps at 43 FR 1227 because there are no stability differences between povidone-iodine and poloxamer-iodine complexes. The comment pointed out that both complexes are currently being marketed as stable products in synthetic soap formulations and argued, therefore, that poloxamer-iodine complex should be made Category III, as was povidone-iodine complex.

The statement regarding incompatibility of iodophors, such as povidone-iodine, that appeared at 43 FR 1236 was in error. The response to comment 70 (43 FR 1221) was correct in stating that povidone-iodine can be formulated in soaps without incompatibility problems. In addition, the list of Category III active ingredients at 43 FR 1229 correctly listed povidone-iodine as a Category III antimicrobial soap. The agency recognizes that both povidone-iodine and poloxamer-iodine complexes can be formulated in soaps without encountering stability problems and will address soap formulations of both complexes in a future issue of the **Federal Register**. (See comment 19.)

J. Comments on Quaternary Ammonium Compounds (quats)

44. One comment requested that benzalkonium chloride be placed in Category I as a skin antiseptic, a patient preoperative skin preparation, and a skin wound protectant, in addition to its present Category I classification as a skin wound cleanser. In support of its request, the comment cited several surgery textbooks and other references that recommend use of benzalkonium chloride at concentrations ranging from 1:750 to 1:5,000 as a preoperative skin preparation, surgical scrub, skin antiseptic for venipuncture, and in urinary tract procedures, especially in catheterized patients (Ref. 1). The comment also submitted 2 studies on a product containing benzalkonium chloride at a concentration of 1:1,000: (1) an in vitro study to demonstrate that this product formulation acts as a physical chemical barrier against contamination by microorganisms, and (2) a study on induced wounds on the arms of 10 healthy subjects to present evidence that this product is nonirritating and neither delays healing nor favors the growth of microorganisms (Ref. 2).

In the previous tentative final monograph, a 1:750 (0.13 percent) use concentration of benzalkonium chloride was proposed as a Category I "skin wound cleanser" (43 FR 1220 at 1236 to 1237). However, this concentration of benzalkonium chloride was categorized as Category III for other uses requested by the comment, i.e., "skin antiseptic," "skin wound protectant," and "patient preoperative skin preparation." The agency stated that it was "not seriously concerned with the safety of 'quats' for 'first-aid' uses, i.e., in skin wound cleansers, skin wound protectants, and skin antiseptics" (43 FR 1236). The agency also stated that "before 'quats' in general can be finally classified for such uses, the following minor issues must be resolved: delay of skin wound repair, contact dermatitis, and sensitivity to 'quats' " (43 FR 1236). The agency limited the use concentration to not greater than 1:750 and advised that data are needed to establish the minimum and maximum concentrations to be included in the monograph. (See comment 53, 43 FR 1219, and 43 FR 1236 to 1237.)

In this amended tentative final monograph for first aid antiseptic drug products, the agency is combining the former categories "skin wound cleaner," "skin wound protectant," and "skin antiseptic" into a new category "first aid antiseptic." (See comment 13.) The other uses for benzalkonium chloride requested by the comment, e.g., "patient preoperative skin preparation" will be addressed in the segment of this rulemaking dealing with uses other than first aid in a future issue of the **Federal Register**.

The agency has evaluated the scientific review of published articles (Ref. 1), as well as data from safety and effectiveness studies on a product containing benzalkonium chloride (Ref. 2). In the studies of the benzalkonium chloride product (1:1000 (0.10 percent)), uniform superficial wounds were made by the ammonium hydroxide blister method on the forearms of each of 10 human test subjects. Tests wounds were treated three times daily for 3 days with benzalkonium chloride and occluded. The control site was untreated and occluded. Quantitative evaluations of resident skin bacteria recovered from test wound and control wound sites demonstrated that benzalkonium chloride significantly reduced the resident bacteria (i.e., 1 $\log_{10}$). In addition, the study showed that although the treated wounds showed a greater degree of erythema than the untreated wounds on observation days 3 and 5, no other significant differences were observed for crust/scab formation, erythema, or epithelization. The agency believes that these data show that benzalkonium chloride is nonirritating and does not interfere with healing of minor wounds.

Based on the new data, the agency concludes that the safe and effective range for benzalkonium chloride has been established between 0.1 percent to 0.13 percent. Because the concerns that the agency raised in the previous tentative final monograph have now been satisfactorily resolved, the agency is including benzalkonium chloride (1:1000, 0.1 percent to 1:750, 0.13 percent) in this tentative final monograph for first aid antiseptic drug products.

References

(1) "Benzalkonium chloride (Zephiran)," unpublished report submitted by Sterling Drug, Inc., Comment No. C00116, Docket No. 75N–0183, Dockets Management Branch.

(2) Unpublished Clinical Wound Healing Studies on Medi-Quik® submitted by Sterling Drug, Inc., Comment No. SUP013, Docket No. 75N–0183, Dockets Management Branch.

45. One comment objected to the 1:750 use concentration limit established for quaternary ammonium compounds in proposed § 333.40(a). The comment submitted safety and efficacy data for a product with a 1:200 concentration of methylbenzethonium chloride, a quaternary ammonium compound. The comment stated that these studies, as well as previously submitted data and a long history of marketing use for this product, demonstrate that this ingredient is safe for consumer use at this concentration. The comment contended that the 1:750 use concentration limit selected by FDA is completely arbitrary and requested that a Category I skin wound cleanser classification be given to this 1:200 concentration of methylbenzethonium chloride.

As discussed in comment 13, skin wound cleansers have been incorporated into a broader group of antimicrobial containing drug products that are designated as first aid antiseptics in this tentative final monograph. The agency has reviewed several reports previously reviewed by the Panel and new material submitted since the Panel report was published, including the data submitted with this comment in the context of this new category. Based on a review of these data, the agency concludes that the 1:200 (0.5 percent) concentration of methylbenzethonium chloride is safe and effective for first aid use on minor cuts, scrapes, and burns.

Apparently, in its original evaluation of these data on methylbenzethonium chloride, the Panel overlooked the 1:200 concentration and referred only to the usual marketed 1:750 concentration of benzalkonium chloride (0.13 percent) as the standard for all quaternary ammonium compounds. The data show that a 0.5-percent concentration of methylbenzethonium chloride is safe and nonirritating. Human studies by Killian (Ref. 1) and by Withers and Hale (Ref. 2) utilizing 0.5 percent methylbenzethonium chloride indicated that the product did not show any significant primary skin irritation, skin fatigue, or sensitization. A study by Vignec (Ref. 3) provided further support that the drug is safe and nonirritating. Vignec used 0.5 percent methylbenzethonium chloride solution full strength on 138 infants suffering from diaper irritation, minor skin conditions, and excoriation and concluded that the drug was safe and nonirritating. The fact that the solution of methylbenzethonium chloride was used under the occlusion of a diaper without evidence of irritation strongly suggests the safety of this concentration.

Maibach (Ref. 4) reported that, even after a 21-day application of 0.5 percent methylbenzethonium chloride under occlusion, minimal irritation was observed. In this study, a 2- by 2-centimeter patch of nonwoven fabric impregnated with 0.2 mL of methylbenzethonium chloride solution was applied to each subject's back and occluded with tape. The patch was removed every 24 hours. After the test site was read, a freshly medicated patch was applied to the same area. The cumulative irritation index score for the 0.5 percent methylbenzethonium chloride preparation was 8.19 and 5.50 out of a possible score of 84. A second study by Maibach (Ref. 5) on 200 subjects used the standard Draize human sensitization test. The investigator concluded that there was no evidence of contact sensitization to the product. Therefore, the agency concludes that a concentration range of 1:750 (0.13 percent) to 1:200 (0.5 percent) of methylbenzethonium chloride is safe and nonirritating as a first aid antiseptic.

The agency's detailed comments and evaluation of the data and the references are on file in the Dockets Management Branch (Ref. 6).

References

(1) Killian, J. A., "Summary of Local Irritation Actions on Skin and of Sensitizing Properties of Bactine," Section II–A, unpublished report to Miles Laboratories, Inc., 1949, OTC Vol. 020088.

(2) Withers, O. R., and R. Hale, "Skin Tests to Bactine on Hypersensitive Patients," unpublished report to Miles Laboratories, Inc., 1951, OTC Vol. 020088.

(3) Vignec, A. J., "Treatment of Diaper Rash," unpublished report to Miles Laboratories, Inc., 1952, OTC Vol. 020088.

(4) Maibach, H. I., "21-Day Cumulative Irritancy Assay," unpublished report, Miles Medical Department Study No. 2213, 1977, Exhibit 6 of SUP014, Docket No. 75N–0183, Dockets Management Branch.

(5) Maibach, H. I., "Modified Draize Human Sensitization—200 Subjects," unpublished report, Miles Medical Department Study No. 2226, 1978, Exhibit 8 of SUP014, Docket No. 75N–0183, Dockets Management Branch.

(6) Letter from W. E. Gilbertson, FDA, to E. B. Peel, Miles Laboratories, Inc., coded LET 038, Docket No. 75N–0183, Dockets Management Branch.

K. Comment on Tribromsalan

46. One comment stated that tribromsalan in its commercially pure form is not a photosensitizer and submitted an unpublished study to support its contention (Ref. 1). In the study the test agent was applied to 25 subjects for 24 hours, followed by exposure to three Minimal Erythema Doses of solar-simulated radiation twice weekly for 3 weeks. The subjects were challenged 10 to 14 days after the last exposure, and the reactions were evaluated 48 and 72 hours later. The study results demonstrated that pure tribromsalan did not cause photocontact allergy, whereas tribromsalan containing 45 percent dibromsalan did. The investigators speculated that the photosensitizing potential attributed to tribromsalan is caused by the presence of dibrominated contaminants.

In addition, the comment included a statement from an expert who had testified before the Panel. This expert had maintained that photosensitization attributable to tribromsalan had not occurred recently and that any earlier cases attributed to tribromsalan were probably due to cross-reactions in patients sensitized to ingredients such as dibromsalan, bithionol, and tetrachlorosalicylanilide or, less likely, hexachlorophene or dichlorophen. The comment requested that tribromsalan be removed from Category II status.

In the **Federal Register** of October 30, 1975 (40 FR 50527), FDA issued a final regulation (21 CFR 310.508) declaring any drug product containing certain halogenated salicylanilides (including tribromsalan) to be a new drug, stating that these ingredients are not generally recognized as safe and effective for use as active or inactive ingredients in any drug product. The study submitted with the comment does not contain sufficient new information to allow the agency to consider tribromsalan generally recognized as safe and effective for OTC drug use. The submitted study, which has since been published (Ref. 2), showed that 2 of the 25 subjects became photosensitized with the sample of tribromsalan containing 45 percent dibromsalan, whereas no subjects had a reaction to the more purified sample of tribromsalan. One of the two subjects who was photosensitized to the tribromsalan that contained dibromsalan developed cross-reactions to the purer sample of tribromsalan. Five subjects who were sensitized by tetrachlorosalicylanilide also showed cross-reactivity to the purer sample of tribromsalan.

When the regulation was published in 1975, the agency recognized that manufacturing limitations for tribromsalan used in earlier formulations resulted in contamination with higher concentrations of more potent photosensitizing chemical impurities, such as dibromsalan and metabromsalan. The agency also noted that the level of impurities was reduced with improved manufacturing techniques and that tribromsalan sensitization was declining, but had not disappeared. In addition to the problem of photosensitization, the agency was concerned about the lack of toxicological data adequate to establish a safe level for use. Another concern was the adverse benefit-to-risk ratio. (See 40 FR 50527 at 50528 and 50530.) In the absence of adequate data to answer these concerns, the provisions of the regulation in § 310.508 for tribromsalan remain in effect. A new drug application containing appropriate toxicological and manufacturing controls information may be submitted to obtain marketing approval for any product containing tribromsalan.

References

(1) Kaidbey, K. H., and Kligman, A. M., "The Photomaximization Test for Identifying Photoallergic Contact Sensitizers," unpublished study, Comment No. C00095, Docket No. 75N–0183, Dockets Management Branch.

(2) Kaidbey, K. H., and Kligman, A. M., "The Photomaximization Test for Identifying Photoallergic Contact Sensitizers," Contact Dermatitis, 6:161–169, 1980.

L. Comments on Triclocarban

47. Several comments requested Category I status for triclocarban as an active ingredient in antimicrobial soaps and presented new safety data. These data included information to elucidate the metabolic pathways and the pharmacokinetics of triclocarban, short-term toxicity data in animals to determine the target organ for toxicity

and the effect and no-effect levels of use (Ref. 1), long-term toxicity data in animals (Ref. 2), and metabolism data in neonate monkeys (Ref. 3). The comments argued that the data confirmed historical experience showing that triclocarban can be safely used in soaps by infants and adults.

The agency has evaluated data and information submitted by the comments and advised a manufacturer that the study entitled "Twenty-Four Month Dietary Toxicity/Carcinogenicity Study of TCC in Rats" (Ref. 2) served to resolve the agency's safety concern regarding blood levels, target organ toxicity, and no effect levels (43 FR 1210 at 1233) and that triclocarban can be recognized as safe for OTC daily topical use in a concentration of 1.5 percent (Ref. 4). However, as stated in comments 10 and 26, antimicrobial soaps making only cosmetic claims are no longer being considered in this rulemaking.

In the previous tentative final monograph, triclocarban (1.5 percent) was categorized in Category III as a skin wound cleanser, and in Category II as a skin antiseptic and skin wound protectant. However, as discussed in comment 13, the agency is no longer using the product category designations of skin antiseptics, skin wound protectants, and skin wound cleansers. Instead, those product categories have been combined into a first aid antiseptic category. The agency has reassessed data that were discussed in the Panel's report (Refs. 1 and 3, 39 FR 33103 at 33125) in light of the first aid antiseptic category, and is proposing a Category III classification for effectiveness for triclocarban (1.5 percent) not in soap forms for use as a first aid antiseptic.

References

(1) Comments No. SUP018, C00099, C00115, and CP0002, Docket No. 75N–0183, Dockets Management Branch.

(2) Comments No. SUP041 and CP0004, Docket No. 75N–0183, Dockets Management Branch.

(3) Comments No. MM0005 and LET047, Docket No. 75N–0183, Dockets Management Branch.

(4) Letter from W. E. Gilbertson, FDA, to G. Roush, Jr., Monsanto Co., coded LET032, Docket No. 75N–0183, Dockets Management Branch.

M. Comments on Triclosan

48. A number of comments submitted data and information from microbiological, mutagenicity, metabolism, cross-sensitization, photosensitization, and drug experience studies on triclosan (Ref. 1). The comments stated that the data and information show that triclosan (up to 1.0 percent) is safe and effective and that triclosan should be placed in Category I for use in the categories that were defined in the previous tentative final monograph, i.e., skin antiseptic, skin wound cleanser, skin wound protectant, antimicrobial soap, health-care personnel handwash, patient preoperative skin preparation, and surgical hand scrub. In addition, one comment submitted information on triclosan (0.1 percent) for the treatment of diaper rash and on triclosan (0.1 percent) combined with benzocaine for the treatment of sunburn (Ref. 2).

One comment from the manufacturer of triclosan objected to the agency's expressed concern, as stated in the tentative final monograph (43 FR 1210 at 1231 and 1233), that there is proliferation of products containing triclosan marketed to the American consumer (Ref. 3). Arguing that the agency's concerns were without factual basis, the comment submitted sales data, held confidential under 21 CFR 10.20(j)(2)(i)(*d*), showing that overall sales of triclosan in the United States have in fact decreased from 1973 to 1977 and that sales for use in bar soaps and deodorants have also declined from 1973 to 1977. The comment pointed out that it has exclusive United States patent rights for triclosan and that no license has been, or will be, granted under these patents. The comment added that to the best of its knowledge triclosan is not used in infant clothing, a use mentioned in the tentative final monograph (43 FR 1231). The comment stated that if triclosan is placed in Cateqory I for use in antimicrobial soaps, it would limit sales of triclosan to OTC use in antimicrobial and deodorant soaps, underarm deodorants, and registered Environmental Protection Agency (EPA) pesticide products. In the future, sales might be extended to include approved new drug applications. The comment also pointed out that the statement at 43 FR 1233 about the EPA's Office of Special Pesticide Review preparing a report on the proliferation of triclosan-containing products is in error, and that the erroneous statement apparently resulted from a miscommunication between FDA and EPA staff. The comment concluded that the concerns about proliferation raised by the agency in the tentative final monograph should not prevent triclosan from being placed in Category I.

Another comment from the manufacturer of triclosan submitted validation reports and raw data from a 2-year chronic oral toxicity study in rats, and carcinogenicity and reproduction studies conducted in mice, rats, rabbits, and monkeys by Industrial Bio-Test Laboratories (IBT) (Refs. 4, 5, and 6) and asserted that its validation of the studies shows that triclosan is safe.

Several comments objected to the agency's restriction that antimicrobial soaps containing triclosan can only be formulated in a bar soap to be used with water (43 FR 1210 at 1229) (Ref. 1). The comments argued that such a restriction was not applied to the other Category III uses of triclosan, i.e., skin antiseptic, skin wound cleanser, and skin wound protectant, and that such a restriction was not recommended by the Panel in the advance notice of proposed rulemaking. The comments suggested that the footnote under "antimicrobial soaps" limiting triclosan to bar soap was probably intended to apply to cloflucarban, which, like triclocarban, is known for its "physical and/or chemical incompatibility."

With regard to safety, the agency evaluated the validation reports to support long-term use of the ingredient (Refs. 4, 5, and 6) and advised the manufacturer of triclosan that the IBT studies were invalid because of numerous problems. The agency's detailed comments and evaluations on the data are on file in the Dockets Management Branch (Ref. 7).

The manufacturer subsequently stated its intent to no longer rely on the 2-year chronic oral toxicity IBT study (Ref. 8), and submitted a final report from a new 2-year chronic oral toxicity study in rats (Ref. 9). Pending completion of the agency's evaluation of this new 2-year study, triclosan remains classified in Category III for safety for long-term use.

The agency has evaluated other data and information (Ref. 1) and advised the same manufacturer that these studies resolved the agency's safety concerns for short-term use of triclosan when used in concentrations up to 1.0 percent, but that additional effectiveness data were needed before the ingredient could be placed in Category I. The agency's detailed comments are on file in the Dockets Management Branch (Ref. 10). In a response to the agency, the manufacturer of triclosan requested further guidance, included effectiveness data from in vivo studies for chronic uses (i.e., antimicrobial soap, health care personnel handwash, and surgical hand scrub), and requested that in future rulemaking proceedings, triclosan (being bacteriostatic and not bacteriocidal) either be excluded from categorization or designated "not applicable" for short-term uses as a patient preoperative skin preparation, skin antiseptic, skin wound cleanser, and skin wound protectant (Ref. 11).

In view of the new category "first aid antiseptic" and the effectiveness criteria in proposed § 333.10 (see comment 57), the agency is tentatively classifying triclosan as Category III for effectiveness as a first aid antiseptic. The use of triclosan as a health care personnel handwash, patient preoperative skin preparation, and surgical hand scrub and safety for chronic use will be addressed in the non-first aid segment of this rulemaking dealing with uses other than first aid in a future issue of the **Federal Register**. The use of triclosan for the treatment of diaper rash was addressed in the **Federal Register** of June 20, 1990 (55 FR 25246 at 25277). The use of triclosan for the treatment of sunburn will be addressed in another OTC drug rulemaking covering drug products for this use.

The agency has communicated further with EPA and has ascertained that there is no specific report on the proliferation of triclosan (Ref. 12). Regarding exclusive patent rights, the agency advises that these are not among the determining criteria to establish general recognition of safety and effectiveness, and therefore cannot be used in the evaluation. However, having reviewed the new data along with the previously submitted data, the agency concludes that there is no proliferation problem with triclosan.

Finally, the agency did not intend to restrict formulations of triclosan to bar soap. The agency has reviewed the Panel's recommendations and the footnotes in the previous tentative final monograph (43 FR 1210 at 1229) and finds that triclosan under "antimicrobial soaps" was erroneously marked with the reference to the footnote "Category III only when formulated in a bar soap to be used with water."

References

(1) Comments No. CP0001, SUP019, SUP023, C00103, C00109, SUP031, SUP039, and C00134, Docket No. 75N–0183, Dockets Management Branch.

(2) Comment No. SUP020, Docket No. 75N–0183, Dockets Management Branch.

(3) Comment No. OB0015, Docket No. 75N–0183, Dockets Management Branch.

(4) "Two Year Chronic Oral Toxicity Study With Fat 80′ 023/A in Albino Rats," Comment No. C00109, Volume 1, Appendix E, and Comment No. C00139, Volumes 1 through 8, Docket No. 75N–0183, Dockets Management Branch.

(5) "Eighteen Month Carcinogenicity Study with Fat 80′ 023/A in Albino Mice," Comment No. C00109, Volume 3, Appendix I, and Comment No. C00139, Volume 9, Docket No. 75N–0183, Dockets Management Branch.

(6) "Three Phase Reproduction Study Albino Rats and Rabbits, Bacteriostat CH 3565," Comment No. C00134, TAB 7, and Comment No. C00139, Volumes 10 through 11, Docket No. 75N–0183, Dockets Management Branch.

(7) Letter from W. E. Gilbertson, FDA, to R. Bernegger, Ciba-Geigy Corp., coded LET028/ANS, Docket No. 75N–0183, Dockets Management Branch.

(8) Memorandum of Meeting between FDA Staff and Representatives of Ciba-Geigy Corp., Comment No. MM0007, Docket No. 75N–0183, Dockets Management Branch.

(9) "FAT 80′ 023 2-Year Oral Administration in Rats," Volumes XLI, XLII, and XLIII and "Determination of FAT 80′ 023 in Blood and Tissue Samples Taken During a Two-Year Chronic Oral Toxicity/Oncogenicity Study in Albino Rats," Volume XLIV, Comment No. RPT002, Docket No. 75N–0183, Dockets Management Branch.

(10) Letter from W. E. Gilbertson, FDA, to R. Bernegger, Ciba-Geigy Corp., coded LET034, Docket No. 75N–0183, Dockets Management Branch.

(11) Comments No. MM0003 and C00157, Docket No. 75N–0183, Dockets Management Branch.

(12) Letter from A. E. Castillo, EPA, to W. E. Gilbertson, FDA, coded LET033, Docket No. 75N–0183, Dockets Management Branch.

N. Comments on Drug Combinations

49. Several comments objected to the agency's decision not to allow combinations of an antimicrobial ingredient and a nonantimicrobial active ingredient or ingredients. (See comment 44, 43 FR 1210 at 1217.) The comments requested that the monograph provide for combinations of an antimicrobial active ingredient with a nonantimicrobial active ingredient or ingredients provided that the combinations are "labeled for use solely for the concurrent symptoms indicated for the active ingredients." Some of the comments pointed out that such combinations were submitted to the Panel for review, e.g., a combination of chloroxylenol and petrolatum (39 FR 33103 at 33104). One comment contended that it was contradictory for the agency to reject the chloroxylenol-petrolatum combination and at the same time define a skin wound protectant in § 333.3(h) of the tentative final monograph as a product that provides both a physical and chemical barrier to infection of small, cleansed wounds, in as much as nonantimicrobial ingredients appear to be necessary to provide the physical barrier of a skin wound protectant. One comment specifically requested that the combination of a topical antimicrobial ingredient with a topical anesthetic ingredient be included in the monograph, stating that such a combination has long been recognized as an effective method of treatment. Another comment made a similar request regarding the combination of alcohol and a topical anesthetic ingredient.

The agency agrees with the comments that antimicrobial ingredients (including alcohol) to help prevent infection can be combined appropriately with nonantimicrobial ingredients to provide concurrent relief for symptoms of minor cuts, scrapes, or burns provided the combination product meets the requirements of § 330.10(a)(4)(iv) (21 CFR 330.10(a)(4)(iv)).

In the previous tentative final monograph, the agency stated that no combinations of antimicrobial and nonantimicrobial active ingredients "are known to exist" (43 FR 127). The agency's statement was based on the Panel's recommended criteria for combining antimicrobial and nonantimicrobial active ingredients and the Panel's recommendation that "if a skin antiseptic claim is made it must meet the requirement of the definition of a skin antiseptic" (39 FR 33103 at 33106). In accordance with the Panel's criteria, neither the Panel in its report nor the agency in the tentative final monograph recognized any Category I skin antiseptics; therefore, no Category I combinations of skin antiseptic and nonantimicrobial ingredients existed. However, because this tentative final monograph is proposing a new category for first-aid antiseptics instead of the category of skin antiseptics and because the definitions for these categories are different, the agency reviewed the submissions to the Antimicrobial Panel in light of the new definition and has determined that combinations containing first aid antiseptics with a topical anesthetic or with a skin protectant do exist. The agency has tentatively determined that these combinations provide rational concurrent therapy, have been previously marketed OTC, meet the requirements in § 330.10(a)(4)(iv), and can be generally recognized as safe and effective. Accordingly, the agency is including the combinations mentioned by the comment in this tentative final monograph.

The agency proposed in § 348.50(b)(2) of the tentative final monograph for OTC external analgesic drug products (48 FR 5852 at 5868) the following indication for local anesthetics: "For the temporary relief of" (select one of the following: "pain," "itching," or "pain and itching") (which may be followed by: "associated with" (select one or more of the following) "minor burns," "sunburn," "minor cuts," "scrapes," "insect bites," or "minor skin irritation.")

The agency proposed in § 347.50(b)(1) of the tentative final monograph for OTC skin protectant drug products (48

FR 6820 at 6832) the following indication for skin protectants: "For the temporary protection of minor cuts, scrapes, burns, and sunburn." These indications are very similar to the indication for first aid antiseptics in § 333.50(b) of this proposed monograph. Nevertheless, it should be noted that first aid antiseptics are classified in Category I for safety based on labeling that they be indicated for use only on small areas of the body for a minor cut, scrape, or burn and that they have a warning not to apply over large areas of the body. Accordingly, those Category I claims for external analgesic drug products or skin protectant drug products that refer to conditions other than minor wounds, and particularly conditions likely to involve large areas of the body (e.g., sunburn), would be Category II for topical antiseptic-anesthetic and antiseptic-skin protectant combination drug products. Accordingly, the agency is proposing the following combinations as Category I in this tentative final monograph:

(1) Any single first aid antiseptic active ingredient identified in § 333.10 may be combined with any single skin protectant active ingredient identified in § 347.10 provided that the product is labeled according to § 333.60.

(2) Any single first aid antiseptic active ingredient identified in § 333.10 may be combined with any single external analgesic active ingredient identified in § 348.10(a) provided the product is labeled according to § 333.60.

The agency is proposing that these combinations bear the general antiseptic labeling indication that appears in § 333.50(b). In addition, the agency is proposing that these combinations can bear the following additional indications:

(1) Antiseptic-external analgesic combination: "First aid for the temporary relief of" (select one of the following: "pain," "discomfort," "pain or discomfort," or "pain and itching") "in minor cuts, scrapes, and burns."

(2) Antiseptic-skin protectant combination: "First aid for the temporary protection of minor cuts, scrapes, and burns."

50. One comment submitted animal, human, and in vitro studies to establish that a combination of 4.7 percent phenol and 10.8 percent camphor (camphorated phenol) in an oil-based vehicle is safe and effective as a first aid product and skin wound protectant (Refs. 1 and 2). (Camphorated phenol is FDA's preferred common name for complexes of camphor and phenol.) The comment stated that this combination was consistent with the Panel's statement that "when camphor is used with phenol in an oil formulation, the concentration of phenol should be no more than 5 percent" (43 FR 1210 at 1238). The comment further stated that "the clathrate complexing of the two ingredients alters the toxicity materially" and that this product has had a long history of safe use with minimal accidental ingestions. The comment concluded that, because of the product's packaging, there is practically no likelihood of mistaking it for mineral oil or castor oil, as has happened with camphorated oil products. The agency has evaluated the reports submitted by the comment (Refs. 1 and 2) and the data submitted to the Antimicrobial I Panel and has determined that camphorated phenol (containing 4.7 percent phenol and 10.8 percent camphor) is safe and effective for use by consumers as a first aid antiseptic.

In a separate rulemaking for OTC external analgesic drug products, the agency categorized the complex (which was described as a combination in that rulemaking) containing camphor and phenol as Category I for short-term use (i.e., 7 days) as an external analgesic, e.g., "for pain and itching of minor cuts and scrapes." The indication for this drug used as an "external analgesic" (48 FR 5852) is similar to the claims in this proposed first aid monograph.

In the external analgesic rulemaking, the agency proposed the following warning for phenol and phenol-containing products: "Do not apply over large areas of the body or bandage" (48 FR 5852 at 5869). This warning is similar to the warning for phenol proposed by the Antimicrobial I Panel (39 FR 33133) and the agency in the previous tentative final monograph (43 FR 1238): "Warning: Do not * * * cover the treated area with a bandage or dressings." There is also an existing required warning in § 369.20 for carbolic acid (phenol) preparations (more than 0.5 percent) for external use: "Warning—Use according to directions. Do not apply to large areas of the body. If applied to fingers or toes, do not bandage." As discussed in comment 25, the agency has included the warning "Do not use in the eyes or apply over large areas of the body," to the general warnings applicable to all first aid antiseptic drug products. Consistent with the external analgesic tentative final monograph and 21 CFR 369.20, the agency is also proposing a separate warning specific for phenol containing products: "Do not bandage."

As discussed in the external analgesic rulemaking, the agency has verified that the amount of free phenol is reduced when camphor and phenol are combined. The Antimicrobial I Panel stated that "when camphor is used with phenol in an oil formulation, the concentration of phenol should be no more than 5 percent" (39 FR 33103 at 33133). In reviewing data on camphor/phenol combinations, the Antimicrobial I Panel concluded that "The presence of camphor also retards the absorption of phenol after topical application. A 1-hour exposure of the rat tail to a 4.8 percent aqueous phenol solution resulted in the absorption of 71 mg of phenol; whereas, the presence of 10.9 percent camphor combined with 4.5 percent phenol resulted in the absorption of only 16 mg phenol * * *" (39 FR 33122).

The agency concluded in the previous tentative final monograph for OTC topical antimicrobial drug products that "the total concentration of phenol in powders and in aqueous, alcoholic or oil formulations be restricted to less than 1.5 percent. When camphor is used with phenol in an oil formulation, the concentration of phenol should be no more than 5 percent" (43 FR 1210 at 1238). The agency agrees with the comment that, based on the data, the antiseptic phenol combined with camphor can be safely used at a higher concentration than phenol used alone. To reduce the irritating potential of phenol when concentrations of 4.7 percent are used, camphor must be present in excess of that concentration. Accordingly, the agency is including camphorated phenol (containing 4.7 percent phenol combined in a complex with 10.8 percent camphor) in a light mineral oil, U.S.P. vehicle in this first aid antiseptic tentative final monograph.

The agency agrees with the comment that camphor/phenol combinations are unlikely to be mistaken for mineral oil or castor oil and that the adverse reaction information supports the safety of the combination.

References

(1) Comment No. SUP013, Docket No. 75N-0183, Dockets Management Branch.

(2) Comment No. C00116, Docket No. 75N-0183, Dockets Management Branch.

51. One comment from a manufacturer of products containing camphorated metacresol disagreed with the Category II classification of formulations containing more than 5 percent phenol or amyltricresols when used with camphor and with the Category III classification of products containing less than 5 percent phenol or amyltricresols when used with camphor (43 FR 1210 at 1238). The comment claimed that "the special safety and effectiveness of these products is based on the existence of a camphor-metacresol complex or one-phase solution, which acts to release

controlled quantities of 'free' metacresol at completely non-toxic levels." The comment stated that a large number and variety of studies had been conducted to demonstrate the safety, effectiveness, and chemical identity of the complex and that, even though the most modern techniques had not been used, the studies should not have been rejected by FDA. The comment submitted new data (Ref. 1) purporting to show that metacresol has a low toxicity compared with other cresols and phenol; that camphorated metacresol is an effective bactericide; and that the antiseptic action of cresols is not due to protein binding and consequently would not encourage continued release of "free" metacresol from the camphorated metacresol complex. Citing the long marketing history of these products, the comment stated that no adverse drug reactions have been reported. The comment argued that this absence of complaints is especially significant because the products are primarily marketed to doctors, nurses, and paramedics for professional use in industrial settings. These professionals are trained to observe and report adverse reactions, treat a limited clientele, and are in close communication with their pharmaceutical suppliers. The comment requested, for the above reasons, that products containing the combination of camphorated metacresol be reclassified into Category I for safety and effectiveness for use as a skin wound cleanser and skin wound protectant without restriction on the metacresol content.

The agency has evaluated the data and concludes that camphorated metacresol limited to a range of camphor 3 to 10.8 percent and metacresol 1 to 3.6 percent in a 3:1 ratio is safe and effective as a first aid antiseptic.

Subsequent to the previous tentative final monograph, the recommendations on camphorated metacresol made by the Advisory Review Panel on OTC Antimicrobial II Drug Products in conjunction with its review of OTC antifungal drug products were published in the **Federal Register** of March 23, 1982 (47 FR 12480 at 12536). That Panel reviewed cresol, the mixture of ortho, meta, and para cresol, and concluded that "Cresol is structurally and pharmacologically related to phenol and * * * is more active against bacteria than phenol and has a phenol coefficient of 2 to 3. The three chemical isomers of cresol (*m*-cresol, *o*-cresol, *p*-cresol) vary little in bactericidal properties" (47 FR 12536). The agency agrees with these findings.

In a separate rulemaking for OTC external analgesic drug products, the agency regarded metacresol as similar to phenol and categorized camphorated metacresol as Category I for short-term use (i.e., 7 days) as an external analgesic, e.g., for pain and itching of minor cuts and scrapes (48 FR 5852 at 5858). This external analgesic indication is similar to the claims in this proposed first aid monograph. As discussed in the external analgesic rulemaking (48 FR 5858), the agency has determined that metacresol behaves similarly to phenol with respect to bonding with camphor and therefore can be considered a "complex" and categorized as camphorated metacresol.

Based on the available information, which includes recognition of the combination of phenol and camphor as Category I, data showing that metacresol has the same toxicity as phenol or is less toxic, and the new data showing that metacresol bonds to camphor similarly to phenol, the agency has tentatively concluded that camphorated metacresol is Category I when prepared from camphor and metacresol combined in a 3-to-1 ratio not to exceed a concentration of 10.8 percent camphor. Based on a 3-to-1 ratio of camphor to metacresol with a limit of 10.8 percent camphor, the upper limit for metacresol is 3.6 percent. This 3-to-1 ratio results in reduced irritation. The agency is proposing a lower limit of 1 percent metacresol based on information on marketed products submitted by the comment.

In addition, the same warning, "Do not bandage," as discussed in comment 50 with regard to phenol/camphor, will apply to camphorated metacresol.

The comment did not provide sufficient data to establish general recognition of safety of a concentration of metacresol greater than 3.6 percent when this ingredient is combined with camphor. The studies submitted by the comment (Ref. 1) were very limited in scope and were inadequate to demonstrate the safety of higher concentrations. Most of the animal toxicity studies tested only one animal, observed the animal only for a short period of time, and did not include a detailed examination of the animal following drug application. The comment's statements about rate of release of metacresol are unsupported because the comment submitted no information on the quantity of metacresol released under the conditions of use. The comment also did not submit any data to support the safety of concentrations of camphor above 10.8 percent.

The marketing history information submitted in the comment does not provide proof of safety for camphor concentrations above 10.8 percent or metacresol concentrations above 3.6 percent. The safety of camphorated metacresol as a first aid antiseptic above 3.6 percent metacresol and 10.8 percent camphor has not been established.

Therefore, the agency proposes to classify camphorated metacresol (a complex consisting of camphor and metacresol combined in a ratio of 3 parts camphor to 1 part metacresol) at concentrations from 1 to 3.6 percent metacresol and from 3 to 10.8 percent camphor as Category I for use as a first aid antiseptic.

References

(1) Comment No. SUP035, Docket No. 75N–0183, Dockets Management Branch.

(2) Comment No. C00098, Docket No. 75N–0183, Dockets Management Branch.

52. One comment stated that the Panel did not review safety and effectiveness data submitted to it on mercufenol chloride (orthohydroxyphenylmercuric chloride) 0.1 percent and secondary amyltricresols 0.1 percent as single ingredients and in combination for use as a patient preoperative skin preparation, skin antiseptic, and skin wound protectant (Ref. 1). The comment added that the agency did not discuss these ingredients alone or in combination in the previous tentative final monograph.

The comment asserted that secondary amyltricresols, mentioned in the previous tentative final monograph under phenol (43 FR 1210 at 1238), are not equivalent to phenol because of chemical differences and differing antimicrobial properties, formulation concentrations, and patterns of use. The comment requested the agency to make decisions on the safety and effectiveness of this ingredient when used alone, or in combination, as a patient preoperative skin preparation, a skin antiseptic, or a skin wound protectant.

The agency has reviewed the submitted data and finds that they are insufficient to determine the safety and effectiveness of 0.1 percent mercufenol chloride and 0.1 percent secondary amyltricresols either singly or in combination for use as a first aid antiseptic. Another panel, the OTC Miscellaneous External Panel, reviewed data other than that provided in this comment and found mercufenol chloride to be safe for topical use at a 0.056-

percent concentration (47 FR 436 at 441). However, the available data are insufficient to establish the safety of this ingredient at 0.1 percent. Only safety data on animals were submitted by the comment (Ref. 1); in general, these studies were conducted on a very small number of animals, did not detail methodology, and did not adequately describe results (physical conditions of the animals). The submitted in vitro studies also lack sufficient detail to establish the effectiveness of mercufenol chloride.

Secondary amyltricresols are mixtures of isomeric secondary amyltricresols, which are derivatives of phenol, and have pharmacological properties similar to phenol. The agency agrees with the comment that the mixture of secondary amyltricresols is not equivalent to phenol and should be categorized separately from phenol. The submitted safety data included a study by Broom (Ref. 2), who reported that amylmetacresol is relatively nontoxic and less toxic than hexylresorcinol in rats and mice.

No toxicity studies in humans were included in the information provided by the comment. However, in the tentative final monograph for OTC external analgesic drug products, published in the Federal Register of February 8, 1983 (48 FR 5852 at 5858), the agency proposed that metacresol up to a 3.6-percent concentration be considered safe when combined with camphor and that a 3-to-1 ratio of camphor to metacresol reduces the irritating properties of metacresol. Although cresols may cause some irritation when applied to minor wounds, the agency believes that secondary amyltricresols at the concentration requested (0.1 percent) would not present any safety concerns, particularly considering the short-term use of first aid products. The submitted data are, however, inadequate to establish the efficacy of secondary amyltricresols.

Data are also needed to determine the safety and effectiveness of the combination of mercufenol chloride and secondary amyltricresols. Only animal safety data are available, and these studies were limited to determinations of the minimum lethal dose by various routes of administration (Ref. 1). The submitted information on marketing history is not sufficient to provide general recognition of the safety of these ingredients. The data contained isolated reports of the combination of mercufenol chloride and secondary amyltricresols causing occasional skin irritation, such as burning and blistering (Ref. 1), adverse effects that need to be more fully studied.

Most of the effectiveness work on the combination of mercufenol chloride and secondary amyltricresols has been in vitro. The combination is reported to combine the antibacterial activity of the single ingredients, that is, mercufenol chloride, which is primarily active against gram-negative organisms, and secondary amyltricresols, which are primarily active against gram-positive organisms (Ref. 3). One in vivo study on the effectiveness of the combination as a patient preoperative skin preparation showed a substantial reduction in the skin microflora (Ref. 4). However, because neutralizers were not used, bactericidal activity cannot be differentiated from residual bacteriostatic activity. In addition, the effect of the 50-percent alcohol in the alcohol-acetone vehicle was not taken into consideration. Alcohol, 48 to 95 percent, has been classified Category I in this first aid antiseptic rulemaking.

Under the agency's guidelines for OTC drug combination products (Ref. 5), Category I active ingredients from the same therapeutic category that have different mechanisms of action may be combined to treat the same symptoms or condition if the combination meets the combination policy in all respects and the combination is on a benefit-risk basis, equal to or better than each of the active ingredients used alone at its therapeutic dose. Accordingly, both mercufenol chloride and secondary amyltricresols and the combination of these ingredients are placed in Category III. The combination needs further testing of the combined ingredients compared to each individual active ingredient to establish effectiveness of the combination as a topical antiseptic for first aid use.

References

(1) OTC Volume 020093.

(2) Broom, W.A., "A Note on the Toxicity of Amyl-meta-cresol," British Journal of Experimental Pathology, 12:327–331, 1931.

(3) Dunn, C.G., "Germicidal Properties of Phenolic Compounds," Industrial and Engineering Chemistry, 28:609–612, 1936.

(4) Maddock, W.G., and L.K. Georg, "Further Experience with Mercresin," American Journal of Surgery, 45:72–75, 1939.

(5) Food and Drug Administration, "General Guidelines for OTC Drug Combination Products," September 1978, Docket No. 78D–0322, Dockets Management Branch.

53. One comment submitted data on the safety and effectiveness of triclocarban and triclosan combined in a deodorant bar soap and requested that this antibacterial combination in a bar soap be included in the OTC topical antimicrobial final monograph (Ref. 1). The comment mentioned that these data were submitted prior to the publication of the previous tentative final monograph, but were not addressed in that document.

The data were not addressed in the previous tentative final monograph because they were received too late for inclusion in that document. As discussed in comment 26, deodorant bar soaps for which only cosmetic claims are made are considered cosmetics.

Reference

(1) Comments No. LET003 and SUP029, Docket No. 75N–0183, Dockets Management Branch.

54. One comment submitted data on the safety and effectiveness of a product containing a combination of eucalyptol, menthol, methyl salicylate, thymol, and 26.9 percent alcohol for use as a first aid remedy and topical antiseptic for the treatment of minor cuts and scratches (Ref. 1). Noting that the product is marketed primarily as an antiseptic mouthwash, the comment stated that it is also labeled and indicated for the treatment of minor cuts and scratches. The comment added that the safety of the ingredients and the total formulation had been acknowledged by two different FDA advisory panels, i.e., the Cold, Cough, Allergy, Bronchodilator, and Antiasthmatic Panel (Cough-Cold Panel) and the Oral Cavity Panel.

References to antiseptic activity of the individual aromatic oils and their combination in the scientific literature were submitted (Ref. 1). Studies of the individual oils (eucalyptol, menthol, methyl salicylate, and thymol), their vapors, and solutions against a wide variety of gram-positive and gram-negative microorganisms were described in the comment. Phenol coefficients were reported for each of the oils. These coefficients show that each oil is more active than phenol against frequently occurring organisms. For example, the following approximate phenol coefficients have been reported: eucalyptol, 1.8; menthol, 5.1; methyl salicylate, 1.8; and thymol, 27.6 (Ref. 1).

The comment included studies (Ref. 1) to demonstrate that the combination of oils is more effective than each of the individual ingredients and that each of the oils provides a statistically significant contribution to the activity of the product.

Further support for the antiseptic activity of the combination is provided by the in vitro antiseptic activity test proposed by the OTC Oral Cavity Panel. The comment stated that at no time has the product failed to kill all three of the

prescribed microorganisms, *C. albicans, Streptococcus mutans,* and *Actinomyces viscosus,* in less than 5 minutes regardless of the test conditions. This includes tests conducted in the presence of saliva, horse serum, or fetal calf serum, each of which may inactivate certain antiseptic agents.

One submitted clinical study compared the antiseptic effect of the combination product, 70 percent ethanol, and water on the skin flora. The study revealed that a 60-second wash of the skin surface with the combination product results in a statistically significant reduction in numbers of surface bacteria. The comment pointed out that there were no significant differences between the combination product and 70 percent ethanol, a widely recognized and recommended antiseptic agent. A gradual recovery of the bacterial count occurs with time, but significantly reduced counts relative to pretreatment values exist 1 and 3 hours postwash after using the combination product and 70 percent ethanol.

Therefore, the comment requested that the agency consider this combination of ingredients to be Category I as a first aid antiseptic in the antimicrobial monograph.

Data and information on the individual essential oils were reviewed by the Oral Cavity Panel, and these ingredients were categorized as Category I for safety. (See the Federal Register of May 25, 1982, 47 FR 22760.) The agency affirms that Panel's conclusions that these individual essential oils are generally recognized as safe. The Cough-Cold Panel also reviewed the ingredients, except for methyl salicylate, and classified them in Category I for safety (41 FR 38311 at 38312). Methyl salicylate was classified in Category I for safety by the Topical Analgesic Panel (44 FR 69768); this classification was confirmed by the agency in the tentative final monograph for OTC external analgesic drug products (48 FR 5852).

The comment submitted data from in vitro studies showing that a formulation of 0.063 percent thymol, 0.042 percent menthol, 0.055 percent methyl salicylate, and 0.091 percent eucalyptol in 26.9 percent alcohol reduced the number of bacteria in *S. aureus* cultures 5.2 $\log_{10}$ within 1 minute at 37 °C when assayed at 40 percent of the formulation's recommended use concentration. Furthermore, when formulations lacking thymol, menthol, methyl salicylate, or eucalyptol were diluted and assayed as described above, the numbers of bacteria were reduced 0.6, 2.4, 3.1, and 3.4 $\log_{10}$, respectively, thus demonstrating that each essential oil contributed significantly to the total antimicrobial efficacy of the complete formulation. Because concentrations of alcohol exceeding 25 percent (v/v) are necessary to inactivate *S. aureus* within 1 hour (Ref. 2), concentrations of 10.76 percent (v/v), such as that contained in the diluted formulations assayed, would not be expected to have significant antimicrobial activity when tested as a single active ingredient. However, antiseptics prepared as hydroalcoholic tinctures have been demonstrated to be more efficacious than aqueous preparations even when dilutions of the tincture high enough to rule out the bactericidal action of the alcohol are assayed (Ref. 3). Thus, the addition of co-solvents to an aqueous phase can influence antimicrobial activity by either the inherent toxicity of the co-solvent, or through the effect of the co-solvent on the thermodynamic activity of an antimicrobial agent, or both (Ref. 4).

The comment also submitted data from in vivo studies which compared the antimicrobial efficacy of four treatment regimens: a formulation of the above mentioned essential oils in 26.9 percent alcohol; 70 percent (v/v) alcohol; water; and no treatment. Treatment consisted of wiping the skin surface for 1 minute with a 2"×2" sterile gauze sponge soaked in the treatment solution, or the site was left untreated to serve as the nontreated control. Bacterial samples were taken from the skin surface by a contact plate method once prior to treatment, immediately after treatment and again at 1 and 3 hours later. Results of the immediate post-treatment evaluation when compared with pre-treatment bacterial counts showed that 70 percent alcohol, the combination of essential oils in 26.9 percent alcohol, water, and no treatment reduced the numbers of organisms 1.69, 1.51, 0.43, and 0.03 $\log_{10}$, respectively. Statistically significant residual effects were observed at 1 and 3 hours after treatment with 70 percent alcohol and the combination of essential oils in alcohol, while water produced a significant reduction immediately and at 1 hour post-wash. Differences in antimicrobial efficacy between 70 percent alcohol and the combination of essential oils in alcohol at 0, 1, and 3 hours post-treatment were not statistically significant.

Although this combination product contains more than two active ingredients from the same pharmacological group (i.e., eucalyptol, menthol, methyl salicylate, and thymol), paragraph 3 of the agency's "General Guidelines for OTC Drug Combination Products" (Ref. 5) permits such a combination "* * * if the combination offers some advantage over the active ingredients used alone, and the combination is, on a benefit-risk basis, equal to or better than each of the active ingredients used alone at its therapeutic dose." In addition, although the individual ingredients have not been classified, the ingredients may be evaluated as a combination based on paragraph 5 of the agency's "General Guidelines" (Ref. 5), which states that "in some cases an ingredient may be appropriate for use only in a specific combination or data may be available only to support the use of the ingredient in combination but not as a single ingredient. In such cases the ingredient will be placed in Category I for use only in the permissible combinations and not as a single ingredient."

Based on these guidelines and discussion above, the agency believes that the combination of eucalyptol 0.091 percent, menthol 0.042 percent, methyl salicylate 0.055 percent, and thymol 0.063 percent in alcohol 26.9 percent may appropriately be included in this amended tentative final monograph as Category I for first aid antiseptic use.

References

(1) Comment No. C00135, Docket No. 75N-0183, Dockets Management Branch.

(2) Morton, H.E., "The Relationship of Concentrations and Germicidal Efficacy of Ethyl Alcohol," Annals of The New York Academy of Sciences, 53:191–196, 1950.

(3) Dunn, Cecil, G., "Germicidal Properties of Phenolic Compounds," Industrial and Engineering Chemistry, 28:609–612, 1936.

(4) Kostenbauder, H.B., "Physical Factors Influencing the Activity of Antimicrobial Agents," Disinfection, Sterilization and Preservation, Edited by Seymour S. Block, Lea and Febiger, Philadelphia, p. 913, 1977.

(5) Food and Drug Administration, "General Guidelines for OTC Drug Combination Products," September 1978, Docket No. 78D–0322, Dockets Management Branch.

55. One comment requested that the agency consider the combination of epinephrine hydrochloride 0.1 percent and methylbenzethonium chloride 0.25 percent for OTC use in the treatment of minor cuts and abrasions. The comment stated that this combination is rational because it contains an antimicrobial agent, methylbenzethonium chloride, to aid in controlling infections and a vasoconstrictor, epinephrine hydrochloride, to help stop the bleeding of a minor wound. The comment added that epinephrine hydrochloride has been marketed in combination products for 35 years; that its safety and efficacy have been confirmed by the Advisory Review Panel on OTC Hemorrhoidal Drug Products (Hemorrhoidal Panel); and that the agency had classified

methylbenzethonium chloride in Category I as a skin wound cleanser in the tentative final monograph for OTC topical antimicrobial products (43 FR 1210 at 1246).

The agency has reviewed the data submitted by the comment (Ref. 1) and concludes that the data are insufficient to establish the safety and effectiveness of a combination of epinephrine hydrochloride 0.1 percent and methylbenzethonium chloride 0.25 percent to treat minor cuts and scrapes.

As discussed in comment 45, the agency considers methylbenzethonium chloride to be safe and effective as a first aid antiseptic at concentrations of 0.13 to 0.5 percent. Although epinephrine has been used for many years as a vasoconstrictor and bronchodilator, its effect on a skin wound in an area of poor circulation, such as an elderly person's finger or toe, needs further study. It has been suggested that epinephrine should not be applied to an area supplied by end arteries, such as the finger, toe, or ear, because of the danger of vascular insufficiency and sloughing (Ref. 2). It should be determined whether vasoconstriction in such a compromised area could induce gangrene.

The agency also finds the submitted data inadequate to determine the effectiveness of this combination. Epinephrine has been used by injection for many years, particularly in local anesthetics to decrease bleeding during surgical procedures, but it has not been as extensively used topically to treat skin wounds. Most of the studies on human skin cited by the comment used either local or intramuscular injections of epinephrine, and not topical applications. No human skin wound studies using epinephrine to stop bleeding were cited by the comment. Further testing of the combination is necessary to determine its effectiveness as a first aid antiseptic for minor cuts and scrapes.

The agency's detailed comments and evaluations on the data and its recommendations for additional studies are on file in the Dockets Management Branch (Ref. 3).

References

(1) Comment No. C00149, Docket No. 75N–0183, Dockets Management Branch.

(2) Denton, J., R.L. Schreiner, and J. Pearson, "Circumcision Complication," Clinical Pediatrics. 17:285–6, 2978.

(3) Letter from W.E. Gilbertson, FDA to K. Johannes, Plough, Inc., coded LET040, Docket No. 75N–0183, Dockets Management Branch.

O. Comments on Testing

56. Several comments requested that the effectiveness requirements for the skin antiseptic drug product category be similar to the requirements for other antimicrobial categories, such as the patient preoperative skin preparation or surgical scrub, for which effectiveness data must show a reduction of the number of bacteria on the skin, and that studies for demonstrating prevention of overt skin infection not be required.

Several comments submitted protocols for determining the in vitro effectiveness of products for general antiseptic use. The lists of microorganisms to be tested varied; but *P. aeruginosa, S. aureus,* and *E. coli* were included in each protocol because they were considered to be the organisms commonly encountered.

One comment asked that efficacy data be reviewed in light of the relevancy of percentage limits of antiseptic to the label claim and that a minimum limit of antiseptic be established for microbiocidal effectiveness. The comment provided experimental data and described a protocol used to determine the quantitative antimicrobial activity of two products of 10 percent povidone-iodine solution. The protocol specified the test organisms for the microbial suspension and the neutralizer to be used and provided for the addition of organic matter (serum) to the culture media in order to determine the minimal inhibitory concentration of the products at different intervals (i.e., zero hour, 15 seconds, 30 seconds, 1 minute, and 5 minutes). Incubation temperature for this in vitro test was 35 °C.

OTC first aid antiseptic drug products are not intended for the treatment of infection or for the prevention of overt infection, but only as an aid in helping to prevent infection of minor cuts, burns, and scrapes. Therefore, the agency finds that studies demonstrating prevention of overt skin infection, which was included in the previous tentative final monograph as part of the definition for a "skin antiseptic," are not necessary for a first aid antiseptic labeled "to help prevent infection in minor cuts, scrapes, and burns."

Demonstrated in vitro antiseptic bactericidal or bacteriostatic action is of predictive value in projecting in vivo efficacy for first aid antiseptics. Based on the comments and the considerations above, the agency has developed effectiveness criteria and procedures for testing final formulations of first aid antiseptic drug products. As recommended by the comments, the organisms *P. aeruginosa. S. aureus,* and *E. Coli* are identified as organisms to be tested. Neutralizers and culture media are discussed in the testing procedures, which are being proposed for inclusion in § 333.70 of the monograph and are described below.

The agency invites specific comment at this time on the testing requirements being proposed in § 333.70. After reviewing any submitted comments or data, the agency may revise the testing procedures prior to establishing a final monograph. The agency also recognizes that the test procedures may need to be revised periodically as newer techniques are developed and proven adequate.

Therefore, the agency is proposing that an OTC first aid antiseptic drug product in a form suitable for topical application meet the standards of the in vitro test included in § 333.70. Because the agency has received data on hydrogen peroxide topical solution, U.S.P., iodine tincture, U.S.P., and iodine topical solution, U.S.P., sufficient to support efficacy for these drug product formulations (see comments 36 and 39), these drug products, when formulated to meet U.S.P. specifications, are exempt from the in vitro testing procedure described in § 333.70.

57. Two comments requested that the agency clarify its position on final formulation testing of antimicrobial drug products because of apparent contradictions between the response to comment 7 (43 FR 1210 at 1211), statements appearing under the testing guidelines at 43 FR 1240, and the response to comment 90 (43 FR 1224).

The agency agrees that there were some contradictory statements in the previous tentative final rule regarding final formulation testing. The agency clarifies in this amended tentative final monograph that all final formulations are required to meet the specifications in the monograph. The agency has provided a test for effectiveness of OTC first aid antiseptics in § 333.70 of the tentative final monograph (as described in comment 56) to be followed by manufacturers for testing the final formulations of OTC first aid antiseptic drug products. The data are not required to be submitted to FDA by the manufacturer. The agency intends to use the testing procedures set forth in the final monograph for any necessary compliance testing of these products. Products that do not meet the specifications in § 333.70 when tested according to the testing procedures set forth in that section or otherwise approved through the petition process described in § 333.70(f) will be considered in violation of the final regulation.

58. Numerous comments addressed the agency's modifications in the Panel's proposed testing guidelines (43 FR 1239

to 1240), the agency's statements on final formulation testing (43 FR 1211, 1224, and 1240), and specific protocols for upgrading an antimicrobial ingredient from Category III to Category I (43 FR 1242 to 1246). Stating that the testing guidelines were unclear and pointing out inconsistencies between the guidelines and the agency's responses to comments at 43 FR 1211 and 1223 to 1227, a number of comments requested clarification or proposed modifications of a number of items in the guidelines.

Several comments requested specific information or submitted protocols for testing Category III ingredients. One comment requested that manufacturers be permitted to determine which protocol to follow to establish safety or effectiveness of an ingredient. A number of comments objected to the agency's consideration of the testing guidelines as final, and urged revisions in the guidelines for publication in the **Federal Register.**

The agency acknowledges that there were some inconsistencies in the testing guidelines for safety and effectiveness proposed in the previous tentative final rule. The agency does not consider the previous testing guidelines as final. The agency is proposing in this amended tentative final monograph a test for final formulations of first aid antiseptic drug products. (See comment 56 above.) Manufacturers may propose other appropriate testing procedures for inclusion in the monograph, and these will be evaluated by the agency upon request. Suggested safety and effectiveness testing procedures of Category III ingredients not in a final formulation are described in the previous tentative final monograph. (See 43 FR 1240.) Because the agency intends to use the testing procedures set forth in the final monograph (and proposed in § 333.70) for any necessary compliance testing of first aid antiseptic drug products covered by the monograph, manufacturers may also use these procedures to test a formulated product containing a Category III ingredient. The test results could be submitted to the agency as part of the information described in the previous tentative final monograph (43 FR 1240) to support the safety and effectiveness of these ingredients.

59. One comment argued that all requirements for preservative testing and data retention under proposed § 333.65 are outside the scope of the OTC drug review rulemaking procedure and should be deleted from the monograph. The comment pointed out that the agency stated in the tentative final monograph that the present framework of the OTC drug review does not permit a review of inactive ingredients, such as preservatives (43 FR 1218). The comment also stated that preservatives by definition are inactive ingredients (43 FR 1214) and as such are not covered by the monograph. Consequently, the comment concluded it is inconsistent with current policy to retain the requirements in § 333.65 of the monograph. The comment requested that all references to preservative testing be deleted from the monograph, especially because these requirements are already covered by the current good manufacturing practice regulations (21 CFR part 211).

Another comment stated that tests to determine the effectiveness of preservative concentration of antimicrobial ingredients are appropriate. However, this comment, as well as another comment, objected to the data retention requirement in proposed § 333.65(c), pointing out that such a requirement exceeds the agency's inspection authority under the act. The comment stated that "defining regulations for topical antimicrobial products cannot be used as a vehicle for expanding the scope of the statute."

Several comments objected to the definition of antimicrobial preservative under § 333.3(b) and requested that it be modified in the following areas: Limiting the preservative to the minimum effective concentration, the requirement for lack of contribution to the claimed drug effects of the product, and the reference to "inadvertently added microorganisms."

Several comments objected to the modifications of the testing procedures as detailed in § 333.65 (a) and (b) from those in the "U.S.P. Antimicrobial Preservative Effectiveness Test" (Ref. 1) and the "CTFA Preservative Test" (Ref. 2). Stating that various parts of these modifications were incongruous, unclear, and conflicting, the comments requested that the U.S.P. and CTFA tests be retained without modifications.

The agency agrees that preservatives are considered inactive ingredients and, upon further review, concludes that it is not necessary to include preservative testing in the tentative final monograph for antimicrobial drug products. However, preservative ingredients must meet the provisions of 21 CFR 330.1(e). The testing procedures detailed in the "U.S.P. Antimicrobial Preservative Effectiveness Test" (Refs. 1 and 3) and the "CTFA Preservative Test" (Ref. 2) are adequate. Therefore, previously proposed §§ 333.3(b) and 333.65 are not being included in this amended tentative final monograph. FDA encourages drug manufacturers to use the U.S.P. and CTFA tests to assure the adequacy of preservative systems in individual products. In view of this action, it is not necessary to respond to the other comments regarding preservative testing.

References

(1) "United States Pharmacopeia XIX," United States Pharmacopeial Convention, Inc., Rockville, MD, p. 587, 1975.

(2) "Determination of Adequacy of Preservation of Cosmetic and Toiletry Formulations," CTFA Technical Guidelines, The Cosmetic, Toiletry and Fragrance Association, Inc., Washington, DC, 1983.

(3) "United States Pharmacopeia XXII—National Formulary XVII," United States Pharmacopeial Convention, Inc., Rockville, MD, p. 1478, 1989.

II. The Agency's Amended Tentative Final Monograph

A. Summary of Ingredient Categories and Testing of Category II and Category III Conditions

1. Summary of Ingredient Categories

The agency has carefully reviewed the claimed active ingredients submitted to this administrative record (Docket No. 75N–0183) including the advance notice of proposed rulemaking (39 FR 33103) and previous tentative final rule (47 FR 1210) for OTC topical antimicrobial drug products, the advance notice of proposed rulemaking for OTC topical alcohol drug products (47 FR 22324), and the advance notice of proposed rulemaking for OTC topical mercury-containing drug products (47 FR 436). Based upon the proposed definition of a first aid antiseptic discussed in comment 13, the agency has made a tentative classification for first aid antiseptic active ingredients.

In arriving at these classifications, the agency has considered all the available data and information, including an assessment of currently marketed ingredients that are labeled or suggested for use as first aid antiseptics. The concentrations described are based upon submitted data. In each case the ingredient has been extensively marketed and used clinically.

Many of the ingredients included in the tabulation below are in Category II and Category III because of a lack of data on use as a first aid antiseptic. However, all the ingredients have been included, as a convenience to the reader. The agency specifically invites comment and additional data on these ingredients.

The agency published an advance notice of proposed rulemaking for mercury-containing drug products on

January 5, 1982 (47 FR 436). That notice, based upon the recommendations of the Miscellaneous External Panel, proposed to classify OTC mercury-containing drug products for topical antimicrobial use as not generally recognized as safe and effective and as being misbranded. The agency received no comments. The Panel classified the mercurial ingredients, as a group, in Category II; some for lack of safety, some for lack of efficacy, and others due to a lack of both safety and efficacy. However, the Miscellaneous External Panel required bactericidal effect for Category I classification as a topical antimicrobial. Based on the proposed definition of "first aid antiseptic," the agency concludes that ingredients having bactericidal and/or bacteriostatic effects are suitable for inclusion in Category I. The agency's criteria are consistent with the Antimicrobial Panel's definition of an antimicrobial (43 FR 1246), i.e., "A compound or substance that kills microorganisms or prevents or inhibits their growth and reproduction * * *." and with section 201(o) of the act (21 U.S.C. 321(o)), which states: "The representation of a drug, in its labeling, as an antiseptic shall be considered to be a representation that it is a germicide, except in the case of a drug purporting to be, or represented as, an antiseptic for inhibitory use as a wet dressing, ointment, dusting powder, or such other use as involves prolonged contact with the body."

Acceptable first aid antiseptic ingredients must be in appropriate product forms to maintain the necessary prolonged contact with the skin in order to sustain their bacteriostatic action. Adequate bacteriostatic action can be demonstrated through in vitro studies. However, data from in vivo studies, such as the ones described for these products in the previous tentative final monograph (43 FR 1210 at 1242), would also be required for these ingredients to be classified in Category I. In light of these changes, the agency has placed those mercurial ingredients with submitted data, which were formerly in Category II solely for efficacy reasons, into Category III and invites interested persons to comment. These mercurial ingredients include calomel, merbromin, phenylmercuric nitrate, and ortho-hydroxyphenylmercuric chloride (mercufenol chloride). "Mercufenol Chloride" is the established name for "ortho-hydroxyphenylmercuric chloride" as listed in the 1991 edition of the "USAN and the USP dictionary of drug names" (Ref. 1). Mercufenol chloride is also discussed in comment 52.

Reference

(1) "USAN and the USP dictionary of drug names," United States Pharmacopeial Convention, Inc., Rockville, MD, p. 369, 992, s.v. "Mercufenol Chloride."

Poloxamer 188 was included in the previous tentative final monograph as a "skin wound cleanser" (43 FR 1246); but, because this antimicrobial rulemaking contains only ingredients with antimicrobial activity and because poloxamer 188 has no such activity, it is not included in the updated tentative final monograph. Poloxamer 188 may be used as an inactive ingredient or pharmaceutical aid in OTC antimicrobial drug products.

The following list is included as a summary of the categorization of first aid antiseptic active ingredients proposed by the agency.

SUMMARY OF ANTIMICROBIAL ACTIVE INGREDIENTS [1]

Category I
Ingredients generally recognized as safe and effective for OTC first aid use within the established concentration(s)
Single ingredients
Alcohol 48 to 95 percent [2]
Benzalkonium chloride 0.1 to 0.13 percent
Benzethonium chloride 0.1 to 0.2 percent
Hexylresorcinol 0.1 percent
Hydrogen peroxide topical solution U.S.P. [4]
Iodine tincture U.S.P.
Iodine topical solution U.S.P.
Isopropyl alcohol 50 to 91.3 percent [2]
Methylbenzethonium chloride 0.13 to 0.5 percent
Phenol 0.5 to 1.5 percent
Combinations
Eucalyptol 0.091 percent, menthol 0.042 percent, methyl salicylate 0.055 percent, and thymol 0.063 percent in 26.9 percent alcohol [4]
Complexes
Camphorated metacresol (3 to 10.8 percent camphor and 1 to 3.6 percent metacresol) in a ratio of 3:1 [4]
Camphorated phenol (10.8 percent camphor and 4.7 percent phenol) in a light mineral oil, U.S.P. vehicle [4]
Povidone-Iodine complex 5 to 10 percent
Category II
Ingredients not generally recognized as safe for OTC first aid use
Single ingredients
Ammoniated mercury [3]
Cloflucarban
Fluorosalan
Mercuric chloride (Mercury chloride) [3]
Mercuric oxide, yellow [3]
Mercuric salicylate [3]
Mercuric sulfide, red [3]
Mercury [3]
Mercury oleate [3]
Mercury sulfide [3]
Nitromersol [3]
Para-chloromercuriphenol [3]
Thimerosal [3]
Tribromsalan
Vitromersol [3]
Zyloxin [3]
Combinations and/or Complexes
None
Category III
Ingredients for which the available data are insufficient to make a final determination for OTC first aid use [1]
Single Ingredients
Benzyl alcohol [2]
Calomel (mercurous chloride) [3]
Chlorobutanol [2]
Chloroxylenol
Merbromin [3]
Mercufenol chloride (ortho-hydroxyphenylmercuric chloride, ortho-chloromercuriphenol) [3]
Phenylmercuric nitrate [3]
Secondary amyltricresols [4]
Triclocarban
Triclosan
Combinations and/or Complexes
Iodine complex (ammonium ether sulfate and polyoxyethylene sorbitan monolaurate) [4]
Iodine complex (phosphate ester of alkylaryloxy polyoxyethylene glycol)
Mercufenol chloride and secondary amyltricresols [4]
Nonylphenoxypoly (ethyleneoxy) ethanoliodine
Poloxamer-Iodine complex
Triple dye
Undecoylium chloride iodine complex

[1] All ingredients (unless otherwise noted) in Antimicrobial I Drug Products Advance Notice of Proposed Rulemaking (39 FR 33103) and Tentative Final Monograph (47 FR 1210).
[2] Alcohol Drug Products, Advance Notice of Proposed Rulemaking (47 FR 22324).
[3] Mercury-Containing Drug Products, Advance Notice of Proposed Rulemaking (47 FR 436).
[4] Not previously reviewed, but categorized in the amended Tentative Final Monograph.

2. Testing of Category II and Category III Conditions

Recommended testing procedures for evaluating the effectiveness of the complete formulation of a first aid antiseptic drug product are included in proposed § 333.70. Suggested effectiveness testing procedures for active ingredients not in a final formulation and suggested safety testing are described in the previous tentative final monograph (see 43 FR 1210 at 1240 to 1242).

Interested persons may communicate with the agency about the submission of data and information to demonstrate the safety or effectiveness of any topical antiseptic ingredient or condition included in the review by following the procedures outlined in the agency's policy statement published in the

Federal Register of September 29, 1981 (46 FR 47740) and clarified April 1, 1983 (48 FR 14050). That policy statement includes procedures for the submission and review of proposed protocols, agency meetings with industry or other interested persons, and agency communications on submitted test data and other information.

B. Summary of the Agency's Changes in the Panel's Recommendations and in the Agency's Previous Recommendations

FDA has considered the comments and other relevant information and is amending the previous tentative final monograph with the changes described in FDA's responses to the comments above and with other changes described in the summary below. The agency is proposing to amend the regulations for topical antimicrobial drug products for OTC human use by adding Subpart A—First Aid Antiseptic Drug Products to 21 CFR part 333 and by amending § 369.20 (21 CFR 369.20). A summary of the changes made by the agency in this amended tentative final monograph follows.

The agency is proposing that skin wound cleansers and skin wound protectants that contain active antimicrobial ingredients be deleted as separate drug product categories and be included in a new category identified as "first aid antiseptics." (See comment 13.) Ingredients that were classified Category I as skin wound cleansers have been classified in Category I as first aid antiseptics. These are benzalkonium chloride, benzethonium chloride, methyl benzethonium chloride, and hexylresorcinol.

2. The agency is proposing that the drug product category "skin antiseptic" be deleted as a separate category and be included in the drug product category identified as "first aid antiseptics." (See comment 13.)

3. A new statement of identity is proposed for the product categories of skin wound protectants, skin wound cleansers, and skin antiseptics. Products previously in those categories are to be identified as "first aid antiseptics." (See comment 9.)

4. The agency is including the following indication for first aid antiseptics: "First aid to help" (select one of the following: "Prevent," ("decrease" ("the risk of" or "the chance of")), ("reduce" ("the risk of" or "the chance of")), "guard against," or "protect against") (select one of the following: "Infection," "bacterial contamination," or "skin infection") "in minor cuts, scrapes, and burns." (See comment 16.)

Because OTC first aid antiseptics are used for the first aid treatment of minor cuts, scrapes, and burns, as are OTC first aid antibiotics, the agency believes that the indications for these two categories of drugs should be similar. The labeling being proposed for first aid antiseptics in this tentative final monograph, where appropriate, is consistent with the labeling adopted in the final monograph for OTC first aid antibiotic drug products (52 FR 47312). (See 21 CFR 333.150(b).)

With the inclusion of alcohol drug products in this rulemaking, labeling recommended for those products has also been incorporated into the first aid antiseptic labeling proposed in new § 333.50. (See comments 27, 28, 32, and 33.)

5. The agency is proposing the following definition for first aid antiseptics consistent with the indication for that drug product category: "An antiseptic-containing drug product applied topically to the skin to help prevent infection in minor cuts, scrapes, and burns." (See comment 13.)

6. The agency is proposing that skin wound cleansers and skin wound protectants without active antimicrobial ingredients do not fall within the scope of the antimicrobial rulemaking. (See comment 14.) Poloxamer 188 was included in the previous tentative final monograph as a "skin wound cleanser," but is not included in the updated tentative final monograph. This antimicrobial rulemaking will only contain ingredients with antimicrobial activity; poloxamer 188 has no such activity. This will not preclude the use of poloxamer 188 as an inactive ingredient or pharmaceutical aid in OTC antimicrobial drug products.

7. Proposed in vitro testing procedures for testing final formulations for use as first aid antiseptics are included in proposed § 333.70. The results need not be submitted to the agency. However, the agency intends to use these testing procedures for any necessary compliance testing. (See comment 57.)

8. The agency proposes to reclassify several ingredients that were placed in Category III either as skin wound cleansers, skin wound protectants, or skin antiseptics to Category I as first aid antiseptics. These ingredients are iodine (tincture and solution) and phenol (0.5 to 1.5 percent). (See comment 37 on iodine.) Phenol is being reclassified into Category I as a first aid antiseptic because the agency has reevaluated effectiveness data available to the agency from the literature and submissions to the Panel (OTC Volumes 020041, 020042, and 020043) which show that phenol (0.5 percent to 1.5 percent without limitation to its vehicle) meets the proposed effectiveness criteria provided in the definition of a first aid antiseptic.

9. The agency proposes to reclassify povidone-iodine complex and camphorated phenol from Category III as skin antiseptics to Category I as first aid antiseptics. (See comments 38 and 39 on povidone-iodine complex and comment 50 on camphorated phenol.)

10. The agency has placed several ingredients that were not reviewed in the previous tentative final monograph into Category I as first aid antiseptics based on data contained in comments to the previous tentative final monograph and information from other sources. These ingredients are hydrogen peroxide, camphorated metacresol (3 to 10.8 percent camphor and 1 to 3.6 percent metacresol in a ratio of 3 to 1), and a combination product containing eucalyptol 0.091 percent, menthol 0.042 percent, methyl salicylate 0.055 percent, and thymol 0.063 percent in 26.9 percent alcohol. Because chlorhexidine gluconate has never been marketed for use as a first aid antiseptic, it is not being included in the first aid antiseptic rulemaking. (See comments 34 on chlorhexidine gluconate, 36 on hydrogen peroxide, 51 on camphorated metacresol, and 54 on the combination product.)

11. Soaps containing antimicrobial ingredients are considered cosmetics when deodorancy or other cosmetic claims are the only claims made for the product. Deodorant labeling claims for antimicrobial soaps are not included in the amended tentative final monograph. (See comment 10.) Antimicrobial soap as a separate drug product category for first aid use is not being included in the amended tentative final monograph. The use of soaps containing antimicrobial ingredients and labeled for other uses, e.g., health care personnel hand washes, will be discussed in the segment of this rulemaking dealing with uses other than first aid in a future issue of the Federal Register. (See comments 10 and 19.)

12. Based upon the proposed definition of a first aid antiseptic, the agency has revised the labeling to eliminate several indications that were Category I in the previous tentative final monograph. These include "prevents skin infection," "controls infection," "degerming," "kills germs," "bacteriostatic," "bactericidal," "reduces the risk of infection and cross-infection," and "microbiocidal." (See comment 16.)

13. The directions for use are being revised to delete the phrase "after gentle washing with soap and water" because

alkaline soap may be inappropriate for use on damaged tissue. (See comment 20.)

14. The warnings in § 333.92(c)(4) "do not bandage tightly" and in § 333.99(c), which stated "the warning 'Do not use solution with occlusive dressing' may be used instead of the warning 'do not bandage tightly,' " which were proposed for all skin wound cleansers, are not being required for all first aid antiseptic drug products. This includes products containing benzalkonium chloride, benzethonium chloride, and methylbenzethonium chloride. The need for such warnings will be separately evaluated for each ingredient based on the ingredient's sensitizing and irritation potential. (See comment 22.)

15. The warning "Do not bandage" is being required for camphorated metacresol, camphorated phenol, and phenol. (See comments 50 and 51.)

16. The agency proposes to revise the warning "This product is not for use on wild or domestic animal bites. If you have an animal bite, consult your physician immediately." Rather than having the separate warning for animal bites, the agency is proposing to add the term "animal bites" to the warning that lists other conditions that need medical attention. The revised warning is as follows: "In case of deep or puncture wounds, animal bites, or serious burns, consult a doctor." (See comment 23.)

17. The agency proposes to revise and consolidate the warnings for skin wound cleansers, skin wound protectants, and skin antiseptics regarding the length of time these products can be used before consulting a physician. The previous tentative final monograph allowed 10 days of self-medication before consulting a physician. This amended tentative final monograph proposes 7 days for consistency with rulemakings for other topical products. The warning is also being revised so that it does not imply that these products are recommended to treat infection. The warning in § 333.93(c)(5) of the previous tentative final monograph that attempted to describe symptoms of infection to alert consumers when to consult a physician has been included in the new general warning in the amended tentative final monograph. The following warning replaces the separate warnings for the three drug products categories and is proposed for all first aid antiseptics, including alcohol: "Stop use and consult a doctor if the condition persists or gets worse. Do not use longer than 1 week unless directed by a doctor." (See comment 24.)

18. The agency is eliminating several redundant or unnecessary warnings proposed in the previous tentative final monograph. The proposed warning in § 333.93(c)(7), "Do not use on chronic skin conditions such as leg ulcers, diaper rash, or hand eczema," has been deleted because the 1-week use limitation warning and the indication should be sufficient to inform the consumer that first aid antiseptics are not to be used on longstanding skin conditions. The proposed warning in § 333.93(c)(6), "Do not use in the eyes," has been expanded to include "or apply over large areas of the body," which is consistent with the first aid antibiotic tentative final monograph. (See comment 25.)

19. The agency proposes to revise the statement of identity for alcohol drug products proposed by the Miscellaneous External Panel in § 333.98(a) as "alcohol for topical antimicrobial use" to the same statement of identity as other first aid antiseptics, i.e., "first aid antiseptic." (See comment 27.)

20. The agency proposes to delete the warning proposed by the Miscellaneous External Panel in § 333.98(c)(2) for products containing isopropyl alcohol, "Use only in a well-ventilated area; fumes may be toxic." (See comment 32.)

21. The advance notice of proposed rulemaking for alcohol drug products for OTC topical antimicrobial use is being adopted by the agency with changes for clarity, and is being incorporated into this amended tentative final monograph. The lower limit of ethyl alcohol is being reduced to 48 percent because of evidence that 48 percent ethyl alcohol is effective as a first aid antiseptic. (See comment 33.) The indications for alcohol and isopropyl alcohol are being modified for consistency with the other Category I first aid antiseptic ingredients. (See comment 33.)

22. The agency is proposing to change the minimum concentration of povidone-iodine for effectiveness from 7.5 percent to 5 percent because of data from studies on a marketed product showing effectiveness at the lower concentrations. (See comment 39.)

23. The agency is eliminating the requirement that iodophors carry a 2-year expiration date. (See comment 40.)

24. The agency is reclassifying povidone-iodine for first aid antiseptic use to Category I. (See comments 41 and 42.)

25. The agency is proposing to change the upper limit of the concentration for methylbenzethonium chloride to 1:200 (0.5 percent). (See comment 45.) In addition, the agency is proposing to change the upper limit for benthezonium chloride to 1:500 (0.2 percent) based on the recommendation of the Miscellaneous External Panel in its report on OTC drug products for the control of dandruff, seborrheic dermatitis, and psoriasis, published in the Federal Register of December 3, 1982 (47 FR 54646).

26. The agency is removing the proposed restriction that dosage forms of triclosan be formulated only in a bar soap. (See comment 48.)

27. The agency is proposing to allow the combination of a Category I antimicrobial ingredient with a Category I analgesic, anesthetic, or antipruritic ingredient or with a Category I skin protectant ingredient. Therefore, new § 333.20 is being proposed in this amended tentative final monograph to include these combinations. (See comment 49.)

28. The agency is not including previously proposed §§ 333.3(b) and 333.65 in the amended tentative final monograph. Nevertheless, the agency encourages manufacturers to continue to test preservatives according to USP and CTFA tests to assure the adequacy of preservative systems in individual products. (See comment 59.)

29. The term "scrapes" has been substituted for the term "abrasions" in the labeling of the amended tentative final monograph for first aid antiseptics, which is consistent with the first aid antibiotic monograph.

30. In an effort to simplify OTC drug labeling, the agency proposed in a number of tentative final monographs to substitute the word "doctor" for "physician" in OTC drug monographs on the basis that the word "doctor" is more commonly used and better understood by consumers. Based on comments to these proposals, the agency has determined that final monographs and any applicable OTC drug regulations will give manufacturers the option of using either the word "physician" or the word "doctor." This amended tentative final monograph proposes that option. (See § 330.50(e).)

31. Several mercury-containing OTC topical antimicrobials have been reclassified from Category II to Category III for effectiveness. Mercurial ingredients placed in Category II for safety are not being reclassified. The ingredients being reclassified are calomel, merbromin, mercufenol chloride, and phenylmercuric nitrate. (See Part II. A.1.—Summary of Ingredient Categories.) This change is being made in keeping with the revised effectiveness criteria for the drug product category "first aid antiseptic" (see comment 56), which were not available at the time the Miscellaneous External Panel evaluated the effectiveness of mercurial ingredients.

32. The agency is proposing to remove a portion of § 369.20 applicable to OTC first aid antiseptic drug products when the final monograph eventually becomes effective because this portion of the regulations will be superseded by the final monograph (part 333, subpart A; proposed in the Federal Register of July 9, 1982 (47 FR 29986)). The item proposed for removal is the entry for "ANTISEPTICS FOR EXTERNAL USE" in § 369.20.

The agency recognizes that there are other portions of §§ 369.20 and 369.21 applicable to OTC first aid antiseptic drug products that will also be removed eventually, but not necessarily at the time the first aid antiseptic final monograph becomes effective. These items include the entries for "CARBOLIC ACID (PHENOL) PREPARATIONS (MORE THAN 0.5 PERCENT) FOR EXTERNAL USE," "CREOSOTE, CRESOLS, GUAIACOL, AND SIMILAR SUBSTANCES IN PREPARATIONS FOR EXTERNAL USE," and "MERCURY PREPARATIONS FOR EXTERNAL USE" in § 369.20 and the entry for "ALCOHOL RUBBING COMPOUND" in § 369.21. These entries are also applicable to other OTC drug rulemakings and will not be removed until all the applicable rulemakings become final.

The agency has examined the economic consequences of this proposed rulemaking in conjunction with other rules resulting from the OTC drug review. In a notice published in the Federal Register of February 8, 1983 (48 FR 5806), the agency announced the availability of an assessment of these economic impacts. The assessment determined that the combined impacts of all the rules resulting from the OTC drug review do not constitute a major rule according to the criteria established by Executive Order 12291. The agency therefore concludes that no one of these rules, including this proposed rule for OTC first aid antiseptic drug products, is a major rule.

The economic assessment also concluded that the overall OTC drug review was not likely to have a significant economic impact on a substantial number of small entities as defined in the Regulatory Flexibility Act (Pub. L. 96–354). That assessment included a discretionary regulatory flexibility analysis in the event that an individual rule might impose an unusual or disproportionate impact on small entities. However, this particular rulemaking for OTC first aid antiseptic drug products is not expected to pose such an impact on small businesses. Therefore, the agency certifies that this proposed rule, if implemented, will not have a significant economic impact on a substantial number of small entities.

The agency invites public comment regarding any substantial or significant economic impact that this rulemaking would have on OTC first aid antiseptic drug products. Types of impact may include, but are not limited to, costs associated with product testing, relabeling, repackaging, or reformulating. Comments regarding the impact of this rulemaking on OTC first aid antiseptic drug products should be accompanied by appropriate documentation. Because the agency has not previously invited specific comment on the economic impact of the OTC drug review on first aid antiseptic drug products, a period of 180 days from the date of publication of this proposed rulemaking in the Federal Register will be provided for comments on this subject to be developed and submitted. The agency will evaluate any comments and supporting data that are received and will reassess the economic impact of this rulemaking in the preamble to the final rule.

The agency has determined under 21 CFR 25.24(c)(6) that this action is of a type that does not individually or cumulatively have a significant effect on the human environment. Therefore, neither an environmental assessment nor an environmental impact statement is required.

Interested persons may, on or before January 21, 1992, submit to the Dockets Management Branch written comments, objections, or requests for oral hearing before the Commissioner on the proposed regulation. A request for an oral hearing must specify points to be covered and time requested. Written comments on the agency's economic impact determination may be submitted on or before January 21, 1992. Three copies of all comments, objections, and requests are to be submitted, except that individuals may submit one copy. Comments, objections, and requests are to be identified with the docket number found in brackets in the heading of this document and may be accompanied by a supporting memorandum or brief. Comments, objections, and requests may be seen in the office above between 9 a.m. and 4 p.m., Monday through Friday. Any scheduled oral hearing will be announced in the Federal Register.

Interested persons, on or before July 22, 1992, may also submit in writing new data demonstrating the safety and effectiveness of those conditions not classified in Category I. Written comments on the new data may be submitted on or before September 22, 1992. These dates are consistent with the time periods specified in the agency's final rule revising the procedural regulations for reviewing and classifying OTC drugs, published in the Federal Register of September 29, 1981 (46 FR 47730). Three copies of all data and comments on the data are to be submitted, except that individuals may submit one copy, and all data and comments are to be identified with the docket number found in brackets in the heading of this document. Data and comments should be addressed to the Dockets Management Branch (address above). Received data and comments may also be seen in the office above between 9 a.m. and 4 p.m., Monday through Friday.

In establishing a final monograph, the agency will ordinarily consider only data submitted prior to the closing of the administrative record on September 22, 1992. Data submitted after the closing of the administrative record will be reviewed by the agency only after a final monograph is published in the Federal Register, unless the Commissioner finds good cause has been shown that warrants earlier consideration.

List of Subjects

21 CFR Part 333

Labeling, Over-the-counter drugs, Topical antimicrobial drug products.

21 CFR Part 369

Labeling, Medical devices, Over-the-counter drugs.

Therefore, under the Federal Food, Drug, and Cosmetic Act, it is proposed that subchapter D of chapter I of title 21 of the Code of Federal Regulations be amended in parts 333 and 369 as follows:

PART 333—TOPICAL ANTIMICROBIAL DRUG PRODUCTS FOR OVER-THE-COUNTER HUMAN USE

1. The authority citation for 21 CFR part 333 continues to read as follows:

Authority: Secs. 201, 501, 502, 503, 505, 510, 701 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 321, 351, 352, 353, 355, 360, 371).

2. New subpart A, consisting of §§ 333.1 through 333.70, is added to read as follows:

Subpart A—First Aid Antiseptic Drug Products

Sec.
333.1 Scope.
333.3 Definitions.
333.10 First aid antiseptic active ingredients.

Sec.

Subpart A—First Aid Antiseptic Drug Products

§ 333.1 Scope.

(a) An over-the-counter first aid antiseptic drug product in a form suitable for topical administration is generally recognized as safe and effective and is not misbranded if it meets each condition in this subpart and each general condition established in § 330.1 of this chapter.

(b) References in this subpart to regulatory sections of the Code of Federal Regulations are to chapter I of title 21 unless otherwise noted.

§ 333.3 Definitions.

As used in this subpart:

(a) *Antiseptic drug.* In accordance with section 201(o) of the Federal Food, Drug, and Cosmetic Act (the act) (21 U.S.C. 321(o)), "The representation of a drug, in its labeling, as an antiseptic shall be considered to be a representation that it is a germicide, except in the case of a drug purporting to be, or represented as, an antiseptic for inhibitory use as a wet dressing, ointment, dusting powder, or such other use as involves prolonged contact with the body."

(b) *First aid antiseptic.* An antiseptic-containing drug product applied topically to the skin to help prevent infection in minor cuts, scrapes, and burns.

§ 333.10 First aid antiseptic active ingredients.

The active ingredient of the product consists of any of the following within the specified concentration established for each ingredient, and the product is labeled according to §§ 333.50 or 333.60:

(a) Alcohol 48 to 95 percent by volume in an aqueous solution denatured according to Bureau of Alcohol, Tobacco and Firearms regulations in 27 CFR part 20.

(b) Alcohol 26.9 percent when combined in accordance with § 333.20(c).

(c) Benzalkonium chloride 0.1 to 0.13 percent.

(d) Benzethonium chloride 0.1 to 0.2 percent.

(e) Camphorated metacresol (camphor 3 to 10.8 percent and metacresol 1 to 3.6 percent in a ratio of 3 parts camphor to 1 part metacresol).

(f) Camphorated phenol (camphor 10.8 percent and phenol 4.7 percent) in a light mineral oil, U.S.P. vehicle.

(g) Eucalyptol 0.091 percent when combined in accordance with § 333.20(c).

(h) Hexylresorcinol 0.1 percent.

(i) Hydrogen peroxide topical solution U.S.P.

(j) Iodine tincture U.S.P.

(k) Iodine topical solution U.S.P.

(l) Isopropyl alcohol 50 to 91.3 percent by volume in an aqueous solution.

(m) Menthol 0.042 percent when combined in accordance with § 333.20(c).

(n) Methylbenzethonium chloride 0.13 to 0.5 percent.

(o) Methyl salicylate 0.055 percent when combined in accordance with § 333.20(c).

(p) Phenol 0.5 to 1.5 percent.

(q) Povidone-iodine 5 to 10 percent.

(r) Thymol 0.063 percent when combined in accordance with § 333.20(c).

§ 333.20 Permitted combinations of active ingredients.

(a) Any single first aid antiseptic active ingredient identified in § 333.10 may be combined with any single external analgesic active ingredient identified in § 348.10(a) of this chapter provided the product is labeled according to § 333.60.

(b) Any single first aid antiseptic active ingredient identified in § 333.10 may be combined with any single skin protectant active ingredient identified in § 347.10 of this chapter provided the product is labeled according to § 333.60.

(c) The ingredients identified in § 333.10 (b), (g), (m), (o), and (r) may be combined provided the product is labeled according to § 333.60.

§ 333.50 Labeling of first aid antiseptic drug products.

(a) *Statement of identity.* The labeling of the product contains the established name of the drug, if any, and identifies the product as a "first aid antiseptic."

(b) *Indications.* The labeling of the product states, under the heading "Indications," the following: "First aid to help" (select one of the following: "prevent," ("decrease" ("the risk of" or "the chance of")), ("reduce" ("the risk of" or "the chance of")), "guard against," or "protect against") (select one of the following: "infection," "bacterial contamination," or "skin infection") "in minor cuts, scrapes, and burns." Other truthful and nonmisleading statements, describing only the indications for use that have been established and listed in this paragraph, may also be used, as provided in § 330.1(c)(2) of this chapter, subject to the provisions of section 502 of the Federal Food, Drug, and Cosmetic Act (the act) relating to misbranding and the prohibition in section 301(d) of the act against the introduction or delivery for introduction into interstate commerce of unapproved new drugs in violation of section 505(a) of the act.

(c) *Warnings.* The labeling of the product contains the following warnings under the heading "Warnings":

(1) *For products containing any ingredient identified in § 333.10.* (i) "For external use only. Do not use in the eyes or apply over large areas of the body. In case of deep or puncture wounds, animal bites, or serious burns, consult a doctor."

(ii) "Stop use and consult a doctor if the condition persists or gets worse. Do not use longer than 1 week unless directed by a doctor."

(2) *For products containing any ingredient identified in § 333.10 (a) and (l).* "Flammable, keep away from fire or flame."

(3) *For products containing any ingredient identified in § 333.10 (e), (f), and (p).* "Do not bandage."

(d) *Directions.* The labeling of the product contains the following statements under the heading "Directions":

(1) "Clean the affected area."

(2) *For products that are ointments, creams, and liquids.* "Apply a small amount of this product on the area 1 to 3 times daily."

(3) *For products labeled for use as a wet compress.* "Bandage lightly. Keep bandage wet with solution."

(4) *For products packaged as sprays.* "Spray a small amount of this product on the area 1 to 3 times daily."

(5) *For products containing any ingredient identified in § 330.10 (a), (b), (c), (d), (g), (h), (i), (j), (k), (l), (m), (n), (o), (q), and (r) of this chapter.* "May be covered with a sterile bandage."

(6) *For products packaged as liquids or sprays.* "If bandaged, let dry first."

(e) The word "physician" may be substituted for the word "doctor" in any of the labeling statements in this section.

§ 333.60 Labeling of permitted combinations of active ingredients.

Statements of identity, indications, warnings, and directions for use, respectively, applicable to each ingredient in the product may be combined to eliminate duplicative words or phrases so that the resulting information is clear and understandable.

(a) *Statement of identity.* For a combination drug product that has an established name, the labeling of the

product states the established name of the combination drug product, followed by the statement of identity for each ingredient in the combination, as established in the statement of identity sections of the applicable over-the-counter (OTC) drug monographs. For a combination drug product that does not have an established name, the labeling of the product states the statement of identity for each ingredient in the combination, as established in the statement of identity sections of the applicable OTC drug monographs.

(b) *Indications.* The labeling of the product states, under the heading "Indications," the indication(s) for each ingredient in the combination, as established in the indications sections of the applicable OTC drug monographs, unless otherwise stated in this paragraph. Other truthful and nonmisleading statements, describing only the indications for use that have been established and listed in this paragraph, may also be used, as provided in § 330.1(c)(2) of this chapter, subject to the provisions of section 502 of the Federal Food, Drug, and Cosmetic Act (the act) relating to misbranding and the prohibition in section 301(d) of the act against the introduction or delivery for introduction into interstate commerce of unapproved new drugs in violation of section 505(a) of the act.

(1) *For permitted combinations identified in § 333.20(a).* In addition to the required indication identified in § 333.50, the labeling of the product may state, under the heading "Indications," the following additional indication: "First aid for the temporary relief of" (select one of the following: "pain," "discomfort," "pain or discomfort," or "pain and itching") "in minor cuts, scrapes, and burns."

(2) *For permitted combinations identified in § 333.20(b).* In addition to the required indication identified in § 333.50, the labeling of the product may state, under the heading "Indications," the following additional indication: "First aid for the temporary protection of minor cuts, scrapes, and burns."

(3) *For permitted combinations identified in § 333.20(c).* The indications in § 333.50 should be used.

(c) *Warnings.* The labeling of the product states, under the heading "Warnings," the warning(s) for each ingredient in the combination, as established in the warnings sections of the applicable OTC drug monographs.

(d) *Directions.* The labeling of the product states, under the heading "Directions," directions that conform to the directions established for each ingredient in the directions sections of the applicable OTC drug monographs, unless otherwise stated below. When the time intervals or age limitations for administration of the individual ingredients differ, the directions for the combination product:

(1) May not contain any dosage that exceeds those established for any individual ingredient in the applicable OTC drug monograph(s), and

(2) May not provide for use by any age group lower than the highest minimum age limit established for any individual ingredient.

§ 333.70 Testing of first aid antiseptic drug products.

A first aid antiseptic drug product in a form suitable for topical application will be recognized as effective if it contains an active ingredient included in § 333.10 and if at its lowest recommended use concentration it decreases the number of bacteria per milliliter in *Staphylococcus aureus* (ATCC No. 6538), *Escherichia coli* (ATCC No. 8739), and *Pseudomonas aeruginosa* (ATCC No. 9027) cultures (available from American Type Culture Collection (ATCC), 12301 Parklawn Dr., Rockville, MD 20852) by 3 $\log_{10}$ within 10 minutes at 32 °C in the presence of 10 percent serum in vitro. Drugs identified in § 333.10 (j), (k), and (l) are exempt from this testing procedure. Furthermore, an antiseptic drug product for inhibitory use as a wet dressing, ointment, dusting powder, or such other use involving prolonged contact with the body, will be recognized as effective if its active ingredient is included in § 333.10 and if a 1:120 dilution of the formulated drug product in growth medium without neutralizers prevents an increase in the number of organisms from an inoculum of 10^8 organisms of the above cultures when incubated at 32 °C for 48 hours. First aid antiseptic drug products that are not exempt from this provision must meet the specified requirements when tested in accordance with the following procedures unless a modification is approved as specified in paragraph (e) of this section.

(a) *Laboratory facilities, equipment, and reagents*—(1) *laboratory facilities.* To prevent the contamination of test microorganism cultures with extraneous microorganisms, perform the test using aseptic techniques in an area as free from contamination as possible. Because test cultures of microorganisms may be adversely affected by exposure to ultraviolet light or chemicals in aerosols, do not test under direct exposure to ultraviolet light or in areas under aerosol treatment. Do environmental tests to assess the suitability of the testing environment frequently enough to assure the validity of test results.

(2) *Equipment.* Use laboratory equipment that is adequate for its intended use. Thoroughly cleanse the equipment after each use to remove any antiseptic residues. Keep the equipment covered when not in use. Sterilize clean glassware intended for holding and transferring the test organisms in a hot air oven at 200 to 220 °C for 2 hours. Use volumetric flasks, pipets, or accurately calibrated diluting devices when diluting standard and sample solutions. Use plastic or glass Petri dishes having dimensions of 20×100 millimeters. Use covers of suitable material.

(3) *Reagents*—(i) *Phenol stock solution.* Prepare a 5-percent weight to volume solution of phenol by the method described in the "Official Methods of Analysis of the Association of Official Analytical Chemists," Kenneth Helrich (ed.), 15th Ed., 1990, pp. 133–134, which is incorporated by reference in accordance with 5 U.S.C. 552(a) and 1 CFR Part 51. Copies are available from the Association of Official Analytical Chemists, 2200 Wilson Blvd., Suite 400, Arlington, VA 22201–3301, or available for inspection at the Office of the Federal Register, 1100 L St. NW., Washington, DC.

(ii) *Serum.* Use inactivated fetal bovine serum without added preservatives and/or anti-infective products.

(b) *Culture media and diluting fluids*—(1) *Ingredients.* Use Soybean-Casein Digest Medium for culture media and diluting fluids that conform to the standards prescribed by "The United States Pharmacopeia XXII/The National Formulary XVII." In lieu of preparing the media from the individual ingredients, the media may be made from dehydrated mixtures which, when reconstituted with distilled water, have the same or equivalent composition as media prepared from individual ingredients. Media prepared from dehydrated mixtures is to have growth-promoting, buffering, and oxygen tension-controlling properties equal to or better than media prepared from individual ingredients. Adjust the pH of each medium with 1 Normal hydrochloric acid or sodium hydroxide before sterilization, if necessary, so that after sterilization the pH will fall within the specified range prescribed by "The United States Pharmacopeia XXII/The National Formulary XVII." Steam sterilize the media in an autoclave at 121 °C for 20 minutes.

(2) *Neutralizers.* When neutralizers are added to culture media and diluting fluid, perform the following tests.

(i) *Neutralizer inactivation of antiseptic test.* Assay the neutralizer

efficacy for the test antiseptic as follows: Prewarm the test antiseptic, culture medium, test culture, and serum to 32 °C by incubating appropriate volumes of all solutions in a water bath at 32 °C for 5 minutes. Mix 0.8 milliliter of antiseptic (for controls use 0.8 milliliter of sterile water) with 9.0 milliliters of culture medium containing an appropriate antiseptic neutralizer followed by the addition of 0.2 milliliter of the test culture in 50 percent serum. Incubate the mixture of cells, serum, antiseptic, and neutralizer at 32 °C for 10 minutes. Remove aliquots, dilute, and assay for surviving bacteria by the plate-count assay method using diluting and plating media containing appropriate neutralizers, if required. Results obtained showing differences greater than 20 percent between test and control cultures indicate that the neutralizer used to inactivate the test antiseptic is ineffective. Reject results obtained from tests employing ineffective neutralization procedures.

(ii) *Neutralizer effect on bacteria viability test.* Test the effect of neutralizers used to inactivate antiseptic active ingredients on cell viability by diluting aliquots of each test organism culture in Medium A (without neutralizer), specified in paragraph (b)(3)(i) of this section, and in the appropriate diluting fluid (neutralizing medium), specified in paragraph (b)(4) of this section. Determine the number of bacteria in aliquots of appropriate dilutions by the plate-count assay method utilizing growth agar medium containing the same neutralizer concentration as the diluting medium. Determine neutralizer effects on cell viability by comparing the relative number of microorganisms growing on Medium B, specified in paragraph (b)(3)(ii) of this section, with and without added neutralizers. Results obtained showing differences greater than 20 percent between cultures diluted in medium with and without neutralizers indicate that, at the concentration utilized, the antiseptic neutralizer alters the determination of viable cells in the test cultures. Reject results obtained from tests in which the neutralizer employed alters the determination of viable cell numbers.

(3) *Culture media*—(i) *Medium A (without neutralizers).* Use soybean-casein digest fluid medium corresponding to that described in paragraph (b) of this section.

(ii) *Medium B.* Soybean-casein digest agar medium. Same as Medium A, except for the addition of 15 grams of agar per liter.

(iii) *Medium C.* Same as diluting fluid 1, except for the addition of 15 grams of agar per liter.

(iv) *Medium D.* Same as diluting fluid 2, except for the addition of 15 grams of agar per liter.

(v) *Medium E.* Same as diluting fluid 3, except for the addition of 15 grams of agar per liter.

(4) *Diluting fluids*—(i) *Diluting fluid 1.* Diluting medium for neutralizing quaternary ammonium and phenolic antiseptic ingredients. Same as Medium A, except for the addition of 5 grams of lecithin and 40 milliliters of polysorbate 20 per liter.

(ii) *Diluting fluid 2.* Diluting medium for neutralizing iodophor antiseptic ingredients. Same as Medium A, except for the addition of 5 grams of sodium thiosulfate per liter.

(iii) *Diluting fluid 3.* Diluting medium for neutralizing mercurial antiseptic ingredients. Same as Medium A, except for the addition of 1 gram of sodium thioglycollate and 2.5 grams of sodium bisulfite per liter.

(c) *Test organisms.* (1) Use cultures of the following microorganisms:

(i) *Staphylococcus aureus* (ATCC No. 6538).

(ii) *Pseudomonas aeruginosa* (ATCC No. 9027).

(iii) *Escherichia coli* (ATCC No. 8739).

(2) *Preparation of suspension.* Maintain stock cultures on Medium B agar slants by monthly transfers. Alternatively, cultures may be lyophylized and stored at −70 °C. Incubate new stock transfers 2 days at 32 °C; then store at 2 to 5 °C. From stock culture, inoculate tubes of Medium A and make at least 4 but less than 30 consecutive daily transfers in Medium A, incubating at 32 °C, before using the culture for testing. Use a 22- to 26-hour culture of organisms grown in Medium A at 32 °C for the test.

(3) *Determination of cell number in broth cultures.* Prepare serial 1:10 dilutions of each culture in Medium A and determine the number of cells per milliliter of culture by the plate-count assay method. Do not use cultures stored at 4 °C for more than 48 hours for assay. Do not use cultures containing less than 10^9 cells per milliliter.

(4) *Plate-count assay.* For each culture to be assayed, pipet 1 milliliter of each prepared dilution into each of two sterile Petri plates. To each plate, add 20 milliliters of sterile Medium B that has been melted and cooled to 45 °C (if neutralizers are required, use the corresponding agar growth medium with the appropriate neutralizer). Mix the sample with the agar by tilting and rotating the plate and allow the contents to solidify at room temperature. Invert the Petri plates and incubate at 32 °C for 48 hours. Following incubation, count the number of developing colonies. Use Petri plates containing between 30 and 300 colonies in calculating the number of bacteria per milliliter of original culture.

(5) *Test organism antiseptic resistance test.* To insure that antiseptic resistance properties of each organism have not altered substantially, determine the resistance to phenol at 20 °C for each organism as described in "Phenol Coefficient Methods" referenced in paragraph (a)(3) of this section.

(i) *Escherichia coli.* A culture of *Escherichia coli* (ATCC No. 8739) is satisfactory for test purposes if it has resistance to phenol at 20 °C at least as follows:

Phenol	5 min	10 min	5 min
1:90 dilution...	+ or 0	+ or 0	0
1:100 dilution.	+	+	+ or 0

(ii) *Pseudomonas aeruginosa.* A culture of *Pseudomonas aeruginosa* (ATCC No. 9027) is satisfactory for test purposes if it has resistance to phenol at 20 °C at least as follows:

Phenol	5 min	10 min	15 min
1:80 dilution...	+ or 0	+ or 0	0
1:90 dilution...	+	+	+

(iii) *Staphylococcus aureus.* A culture of *Staphylococcus aureus* (ATCC No. 6538) is satisfactory for test purposes if it has resistance to phenol at 20 °C at least as follows:

Phenol	5 min	10 min	15 min
1:60 dilution...	+ or 0	+ or 0	0
1:70 dilution...	+ or 0	+	+

(d) *Test procedures*—(1) *Method 1*—(i) *Method validation.* This test is valid only for those antiseptics that are water soluble and/or miscible and that can be neutralized by one of the subculture media specified in paragraphs (b)(3) and (b)(4) of this section or that can be overcome by dilution.

(ii) *Bactericidal assay procedure.* Prewarm all test solutions by incubating appropriate volumes at 32 °C in a water bath for 5 minutes. Pipet 1.0 milliliter of serum, 1.0 milliliter of appropriate bacterial test culture, and 8.0 milliliters of test antiseptic at its recommended use concentration into a medication tube and mix well. Incubate at 32 °C for 10

minutes. Remove triplicate 1-milliliter sample aliquots and dilute in Medium A containing appropriate neutralizers. Determine the number of surviving organisms per milliliter of test culture by the plate-count method using plating media containing appropriate neutralizers, if required.

(iii) *Bacteriostatic assay procedure.* Prewarm all test solutions by incubating appropriate volumes at 32 °C in a water bath for 5 minutes. Pipet 1.0 milliliter of serum, 1.0 milliliter of appropriate bacterial test culture and 8.0 milliliters of test antiseptic at its recommended use concentration into a medication tube and mix well. Pipet 1.0 milliliter aliquots of this test mixture into triplicate medication tubes containing 100 milliliters of Medium A without neutralizers and mix well. Incubate at 32 °C for 48 hours and determine the number of organisms per milliliter of culture by the plate-count method.

(2) [Reserved]

(e) *Test modifications.* The formulation or mode of administration of certain products may require modification of the testing procedures in this section. In addition, alternative assay methods (including automated procedures) employing the same basic chemistry or microbiology as the methods described in this section may be used. Any proposed modification or alternative assay method shall be submitted as a petition under the rules established in § 10.30 of this chapter. The petition should contain data to support the modification or data demonstrating that an alternative assay method provides results of equivalent accuracy. All information submitted will be subject to the disclosure rules in part 20 of this chapter.

PART 369—INTERPRETATIVE STATEMENTS RE WARNINGS ON DRUGS AND DEVICES FOR OVER-THE-COUNTER SALE

3. The authority citation for 21 CFR part 369 continues to read as follows:

Authority: Secs. 201, 301, 501, 502, 503, 505, 506, 507, 701 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 321, 331, 351, 352, 353, 355, 356, 357, 371).

§ 369.20 [Amended]

4. Section 369.20 *Drugs; recommended warning and caution statements* is amended in subpart B by removing the entry for "ANTISEPTICS FOR EXTERNAL USE."

Dated: May 20, 1991.

David A. Kessler,

Commissioner of Food and Drugs.

[FR Doc. 91–17107 Filed 7–19–91; 8:45 am]

BILLING CODE 4160–01–M