Applications of Causal Inference Methodology to Questions of Age-Related Changes in Physical and Cognitive Function in Elderly Populations

by

Thaddeus James Haight

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Epidemiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ira B. Tager, Chair
Professor William J. Jagust
Professor Ronald D. Lee
Professor Mark J. van der Laan

Spring 2010

Abstract

Applications of Causal Inference Methodology to Questions
of Age-Related Changes in Physical and Cognitive Function
in Elderly Populations

by

Thaddeus James Haight

Doctor of Philosophy in Epidemiology

University of California, Berkeley

Professor Ira B. Tager, Chair

Causal inference methodology represents an analytical framework to evaluate and estimate causal effects based on observational data. This framework can be applied to examine and estimate individual contributions of causal factors in epidemiologic studies of aging. These studies are characterized by complex temporal relationships between these different factors, and confounders of these factors and outcomes of interest (i.e. time-dependent confounding). Unbiased estimation of effects under these conditions lies beyond the scope of conventional statistical methods. Causal inference methods allow for examination and unbiased estimation of population-level (i.e., marginal) effects, and can account for time-dependent confounding. Marginal Structural Models (MSMs) can be used to define causal parameters (i.e., marginal effects) of interest. History-adjusted marginal structural models (HAMSMs), a generalization of MSMs, can be used to define and evaluate marginal effects given time-varying covariates.

These models were applied to examine the causal associations between different factors of interest and physical and cognitive functioning outcomes. Estimation procedures were applied (e.g., targeted maximum likelihood estimation, inverse-probability of treatment weights) to address time-dependent confounding in the data and provide unbiased estimates of the effects examined in these models, which would not have been possible with standard statistical methods.

Study 1

Self-reported leisure-time physical activity (LTPA), based on metabolic equivalents, equal to or greater than public-health recommended-levels, and a measure of body composition (lean: fat mass ratio, L/F), estimated from bioelectric impedance using population-specific prediction equations derived from dual x-ray energy absorptiomery, were examined with respect to a measure of physical function that was based on standard self-report questions. In women 55 and older, a one-unit gain in L/F reduced by 65.5 percent (95% CI: 21.8, 87.4) the report of physical limitation at all four surveys of an 8-year study. A similar reduction was not observed in men; however, there was a 3 percent increase in the report of no limitation at any survey. The effect of high levels of LTPA was a reduction in new physical limitation that occurred at the last survey of 36.8 percent (95% CI: 0.0, 0.92) and 52.7 percent (95% CI: 13.5, 91.9) in men and

women, respectively. In summary, higher LTPA appeared to reduce the risk of future functional limitation conditional on the level of functioning established by L/F.

Study 2

The association of lifetime household secondhand smoke exposure (SHS) and risk of incident dementia was examined in 970 participants in the Cardiovascular Health Cognition Study who were never-smokers and were free of clinical cardiovascular disease (CVD), dementia, and mild cognitive impairment at baseline. Given prior studies have found that SHS is associated with increased risk of CVD and that CVD is associated with increased risk of dementia, interactions between SHS and measures of clinical and subclinical CVD on dementia risk were examined as well. Moderate (16-25 years) and high (>25 years) SHS exposure levels were not independently associated with dementia risk; however, subjects with >25 years of SHS exposure and >25% carotid artery stenosis had a three-fold increase (Hazard Ratio, 3.00; 95% Confidence Interval: 1.03, 9.72) in hazard of dementia, compared to subjects with no/low (0-15 years) SHS exposure and ≤ 25% carotid artery stenosis. High lifetime SHS exposure may increase the risk of dementia in elderly with undiagnosed CVD.

Study3

Leisure-time physical activity (LTPA), based on metabolic equivalents, was examined with respect to walking speed (WS) that was based on standard protocol. A measure of body composition (Lean:fat mass ratio, L/F) (see Study 1), was included as a surrogate of metabolic function (e.g., glucose tolerance) and as a causal intermediate of LTPA and WS. In sex-specific analyses, the direct effects of LTPA on WS were estimated from four separate surveys of an 8-year study, and pooled. Stratified analyses examined effect estimates in different subgroups (e.g., diabetics vs. non-diabetics). Mean WS increased (2.394 ft/sec vs. 2.238 ft/sec in women; 2.418 ft/sec vs. 2.278 ft/sec in men) with higher LTPA (i.e., greater or equal vs. less than public health recommended levels) and higher L/F (i.e. > median vs. ≤ median). In women, the direct effect of LTPA was an increase in mean WS for ≤ median L/F (2.316 ft/sec vs. 2.238 ft/sec) and > median L/F (2.394 ft/sec vs. 2.316 ft/sec). Similar results were observed for the men. Results of the stratified analysis did not differ from the overall analysis. These results indicate a marginal level of direct protection of LTPA for WS, but underscore the influence of LTPA on metabolic intermediates that affect lower body-function.

The application of MSMs and HAMSMs in these different studies illustrate their use to examine causal factors associated with cognitive and physical decline. Moreover, estimation procedures were employed to provide accurate estimates of effects based on these models of physical activity and other factors associated with function in the elderly.

Table of Contents

Study Hypotheses

There is a large body of epidemiological data that demonstrates that physical activity in the elderly is associated with a wide-variety of improved health and functional outcomes. However, the statistical methods that have been used to investigate these relations have failed to account for the complex temporal inter-relatedness between past and subsequent levels of physical activity, confounders of physical activity and the health outcomes themselves—i.e., time-dependent confounding. Similar complications occur for the investigation of other factors of interest (e.g., environmental, social, physiologic) as these relate to functional outcomes as well. Failure to account for such confounding leads to biased estimates of the effects under study. Causal inference methods have been developed to examine marginal effects, based on a class of causal statistical models called marginal structural models (MSMs), and to provide estimates of these marginal effects for treatments (exposures) on health outcomes in the presence of time-dependent confounders that are often is observed in longitudinal observational data. Moreover, conventional statistical methods cannot provide marginal (population-level) effect estimates of the interactions between exposures and important time-varying covariates. The class of causal statistical models called history-adjusted marginal structural models (HAMSMs), and estimators of these models, address this issue as well.

In this research, I use MSMs and HAMSMs, and estimators of the models, to test the following hypotheses in two cohorts of elderly subjects:

Hypothesis I

Physical activity levels that are equivalent to or greater than the public health recommended guidelines, based on metabolic equivalents, and lean body mass, as measured by the ratio of lean body mass to fat mass (lean:fat ratio), contribute ***independently*** to improved physical functioning patterns.

Hypothesis II

Lifetime exposure to secondhand smoke (SHS), as measured by years of living with a smoker, is independently associated with increased dementia risk; and, given the growing evidence that vascular disease contributes to the clinical manifestation of dementia, that SHS increases dementia risk in subjects with underlying clinical or subclinical vascular disease.

Hypothesis III

Levels of physical activity that are equivalent to or greater than the public health recommended guidelines *reduce* the decline of walking speed *independently* of lean body mass, as measured by the ratio of lean body mass to fat mass (lean:fat ratio).

A. The relative benefits of physical activity on the maintenance of walking speed are greater for elderly subjects who are older than 75 years, and/or have been diagnosed with diabetes and/or cardiovascular disease.

The application of causal inference methods to examine these hypotheses provide an opportunity to unravel some of the complex, temporal relationships between factors associated with cognitive and physical decline, and will provide more accurate estimates of the function-sparing effects of physical activity in the elderly than have been available until now.

Chapter 1-Background

It is expected that the population of older adults in the U.S. will increase in size over the next 50 years, and will constitute a higher proportion of the overall population than ever before in history (1). It is projected that this group will comprise 20 percent of the U.S. population in 2030, up from its current percentage of 12 percent, and surpass the population of young people under age 20 years (1). The population of the oldest-old, those 85 and older, estimated to be 4.7 million in 2003, is projected to double by 2030, and to double again in 2050 (1). There are several reasons for the current trend with respect to the unprecedented growth in the elderly population: medical advancements, healthier lifestyle and behavioral adaptations, and availability of services that did not exist for a large majority of seniors until more recently. As a result of these changes, it is expected that the average life expectancy of seniors will continue to grow in the current century (2).

These population trends signify a major public health achievement. People are living longer and many are leading very productive, fulfilling lives in their older years. A recent report from the National Institute on Aging (NIA) found that, among adults 65 and older in the U.S., the reported percentage of people with a disability fell from 20% in 1989 to 15% in 1999 (1). Concurrently, reported heart disease, the leading cause of death among older people, as well as other chronic health conditions, also declined (1). Overall, older people today are healthier, better educated, more active physically and have more resources available to them than their counterparts in previous decades (1). Given an active, healthy lifestyle that starts in youth and continues through the adult years increases a person's chances of a healthy, independent life. The MacArthur study examined several factors associated with successful aging and found that social engagement, physical activity, and cognitive pursuits were associated with longer survival, better health, and less disability and impairment(3). Findings from other studies indicate that on the whole these factors are among those that some would say are indicators of successful aging.

Although the health and welfare for a majority of seniors has improved tremendously in recent decades, many seniors still suffer from a number of chronic diseases that are commonly associated with aging: arthritis, hypertension, depression, etc. It has been reported that eighty percent of seniors have at least one chronic condition and 50 percent have at least two (1). There is evidence of the potential reemergence of diseases that were in decline in the aged population. For example, compared to 1988-1994, the proportion of obese men aged 65-74 in 1999-2000

rose from 24.1% to 33.4%, and from 13.2% to 20.4% among men 75 years and older. A similar increase is reported for women over this period (1). Obesity brings associated problems of increased rates of diabetes, cardiovascular disease, and lower physical function (4). It is expected that the prevalence of hypertension, and the prevalence of Alzheimer's Disease (AD), diseases associated with older age, will continue to rise given the higher risk of these diseases in old age and the growth of the senior population (1, 5). These diseases can have debilitating effects that frequently lead to physical disability, loss of independence, and pre-mature mortality.

In the absence of disease, aging involves physical changes that result, in general, in reduced physical vigor and "well-being". Age-related changes in molecular and physiologic systems increasingly undermine individuals ability to respond to the stresses exerted by the environment (6-8). For example, healthy older people experience decline in cardiorespiratory reserve and lean body mass. Studies have shown a 15 percent decrement in peak VO_2 per decade in healthy seniors 50-75 years of age, based on cross-sectional data, and a greater decrement in comparable populations based on longitudinal studies (9-11). The effects of aging on skeletal muscle can not only lead to reduced lean muscle mass, but also to reduced muscle strength in the remaining lean muscle of older subjects (12-15). Given these physiologic alterations, older adults become increasingly compromised with respect to meeting their metabolic and physiologic needs and performance of everyday tasks, and over time are increasingly challenged to maintain their health and to lead active, productive lives(6, 7, 16). Therefore, it is imperative that researchers develop a better understanding of the complex interplay of factors that can affect the shape and timing of aging processes /events, and to better understand the factors that could contribute to a healthier, free-living elderly population.

Numerous studies have investigated factors and mechanisms of population aging, and the implications of these for health and function among the elderly. For example, molecular and epidemiologic studies of aging, have aimed to understand better the mechanisms of aging, and how factors in the aging process influence the realization of various health outcomes---e.g., physical fitness, physical function, cognitive function, mortality (3, 17-22). In particular, epidemiologic studies of aging challenge researchers in terms of the inclusion and proper characterization of factors and mechanisms relative to the study outcomes under assessment. Part of the challenge of examining and understanding these factors and mechanisms is to think about them in the context of aging populations, where change is inherent and marked by a decline across multiple health factors and physiologic systems. Therefore, to investigate effects of interest, the researcher has to account for changes with respect to levels of other factors under surveillance.

Examination of causal mechanisms in the presence of change that characterize studies of aging is complicated by the fact that physiologic systems of subjects on which the studies are based are interconnected (7, 20). Therefore, it one wants to examine the individual contributions of these various physiologic systems on aging, one has to account for the interrelationships these have with other systems which may affect aging in a similar way. For example, if one wants to consider the decline of neurological systems on brain aging, one has consider them in the context of decline in cardiovascular health, which can lead to cognitive disorders. Similarly, if one wants to consider loss in muscle mass and muscular strength, and the individual contributions of these to lower body mobility, one has to consider the neurologic changes, which may co-occur with muscle loss, that can lead to reduced mobility. Examination of higher order forms of these functions—e.g., physical activity, physical fitness---and the different implications these may have for health and function in older populations bring challenges as well. For example, the

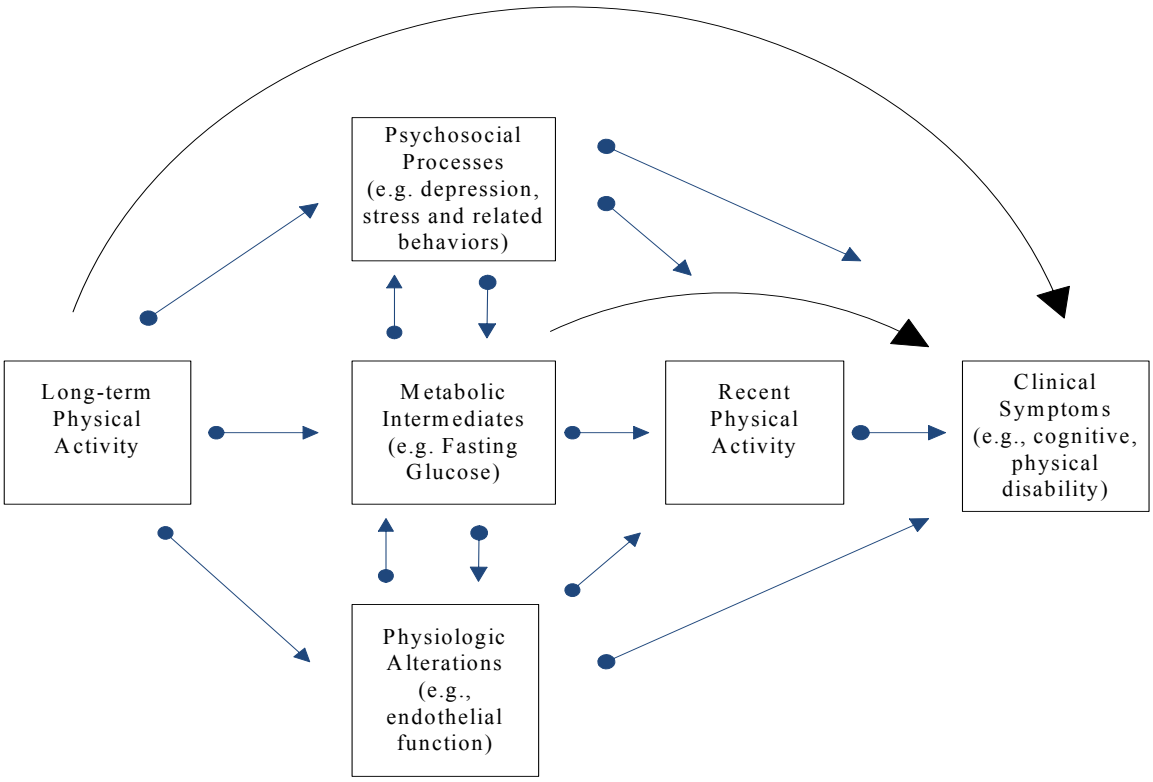
effects of physical activity may indeed be beneficial on subsequent health outcomes, but it may be difficult to identify and estimate these, given the decline in health and function that can accompany aging, which often result in a decline in physical activity levels. Therefore, in an effort to identify and examine the effects of factors that we hypothesize are likely to contribute to subsequent health and functional outcomes, we are likely to encounter concurrent factors which make it difficult to isolate and evaluate the effects of interest.

These challenges, and what is becoming known on a daily basis with regard to the developments in neuroscience, genetics, and the connection between genetic and environmental influences and human health, should give researchers pause with respect to the current understanding of the factors and processes that underlie aging and the age-related outcomes of health, function, and mortality. A more rigorous and comprehensive understanding of these factors and their interconnections is crucial for advancement of the field and the design of interventions for the investigation and potential improvement of health-related outcomes. Current models may not be biologically realistic, nor current methods of analysis satisfactory, to investigate properly the various mechanisms that could lead to improved insights in our assessment of the processes and effects of aging. Improvement in these areas could allow us to better delineate the relative contributions of physical activity and physical fitness, for example, on long-term health, or the effects of these on disease mechanisms in presence of important intermediary factors (e.g., hypertension) known to contribute to poor health outcomes. Also, improved models and techniques of assessment would allow us to better ascertain the reasons that subjects' respond and do not respond to various interventions, and provide more clear explanations for the heterogeneous patterns in physical and cognitive function that are commonly observed in older subjects.

One of the main goals of investigators of population aging has been to better understand the role of physical activity with respect to aging and health. Physical activity is one of the optimal choices to achieve and maintain fitness and health at little or no cost (23). The Surgeon General's 1996 report, and an update of the report in 2007, found that improvement in quality of life could be attained through life-long practice of physical activity and reported that 30 minutes of moderate levels of physical activity daily could reduce the risks of coronary heart disease, hypertension, colon cancer, and diabetes (23, 24). Moreover, physical activity improves cardiorespiratory fitness, muscle strength, and metabolic function (e.g., lipid utilization, glucose control) which not only reduces the risk of adverse health outcomes, but increases physical fitness and cardiovascular reserve (4, 23l, 24-26). Moreover, studies show that individuals who have suffered a health event but who have engaged in physical activity prior or subsequent to their event have a reduced risk of a second event (e.g., 2nd heart attack) compared to their counterparts who were not active before or after their event (3, 4, 27, 28).

Other benefits of physical activity are improved mood, reduce stress, increase self-efficacy, and promotion of healthy behaviors, which are known mediators of health, fitness, and mortality (18, 29, 30). Although the effects of physical activity are considered to be mediated by multiple factors (e.g., lower cholesterol, lower blood pressure, improved cerebral perfusion) known to be responsible for improved health outcomes (e.g., reduced risk of stroke, cardiovascular disease), it has been shown, too, that physical activity may have direct benefits for the molecular and physiologic systems that underlie health and function (e.g., BDNF, reduction of antioxidants, improved mitochondrial function) (31-33).

FIGURE 1. Representation of direct and indirect mechanisms by which physical activity potentially affects various physiologic and psychosocial attributes, and long-term function.



The benefits of physical activity not only occur for the general population, but extend to the elderly and are considered essential to the successful aging of seniors (3, 34, 35). Several studies have demonstrated the health benefits of physical activity for adverse health outcomes such as obesity, impaired physical function, cardiovascular morbidity, and cardiovascular and non-cardiovascular mortality (36-45). The MacArthur Foundation Study, which examined determinants of successful aging (i.e. living without disease), showed that maintenance of physical activity and fitness over one's lifetime reduces the risk of diseases commonly associated with older adults (e.g., heart disease, diabetes, stroke, hypertension) (3). Moreover, this study, as well as others, found that physical activity could offset the health risks associated with smoking, high blood pressure, and coronary artery disease in older adults, and that the protective effects of physical activity extended to those who started exercise later in life (3, 46, 47). Furthermore, physical activity compared to pharmaceutical agents, which are typically administered to elderly patients with medical complications (e.g., hypertension, hyperlipidemia), has the potential to provide health benefits without the side effects and potential harm caused by the use of medications in older adults (e.g., dizziness, cognitive impairment, falls) (48).

Studies also have examined the role of physical activity in cognitive decline and dementia in the elderly, and have shown that older adults who are more physically active, and

have higher levels of cardiorespiratory fitness and improve their levels of physical function, are less likely to experience cognitive decline and dementia (49-56). Moreover, physical activity reduces the behavioral and health risks (e.g., cardiovascular disease, glucose intolerance, smoking, obesity, stress) that are known mediators of poor cognitive health outcomes (e.g., stroke, impaired memory) (57-59) as well as having direct benefits on different brain components (e.g., brain volume, BDNF impact on hippocampus) both in human and animal studies (31, 32, 60).

In addition, evidence that physical activity benefits the elderly comes in the form of a written supplement to the guidelines for physical activity distributed by the Center of Disease Control and the American College of Sports Medicine (CDC/ACSM) for the U.S. adult (over age 18) population (24, 35). These guidelines recommend that the elderly, if possible, should participate in levels of physical activity comparable to those levels recommended for the general US population. In addition, there are guidelines to indicate that seniors should include strengthening exercises as well as engage in activities that increase balance to prevent falls and reduce the risk of fractures that occur with falls (35).

However, it is unclear whether the purported benefits and recommended physical activity guidelines as these apply to seniors have taken sufficient account of the underlying complexities in health and function that characterize these populations. Effects of the aging process vary considerably among individuals, and the differences between older individuals with respect to physical capacity, health, and functional status, compared to non-seniors, is well known (20, 61, 62). For example, diseases, like diabetes, and their relationship to functional status in older subjects are variable and multifactorial (e.g., multiple diseases may synergistically affect function). In addition, factors (e.g., depression) may mediate the relationships between these diseases and their effects on subsequent health outcomes (48). Verbrugge and Jette proposed a conceptual model (i.e., "the disablement process") to account for the various disease pathways linking healthy functioning with physical limitation and disability in older adults (63). Differences in aging and the pathways by which individuals succumb to disease and disability, as illustrated by this model, raise questions with regard to effectiveness of physical activity to offset decline and improve outcomes with respect to the health and functional changes that occur in the elderly. For example, prescription of one set of recommendations for physical activity to seniors as a whole is equivalent to prescription of the same medication and dosage to every senior regardless of these differences. To some extent, the recommendations prescribed by the CDC for adults 65 and older, take into account differences in individuals underlying capacity. However, the guidelines that are offered prescribe activity levels ***relative*** to individuals' capacity (35). Any benefits conferred by the levels of physical activity based on these recommendations would appear to vary considerably and to depend on one's underlying capacity for physical activity; thus, the guidelines are rather limited and lack explicitness with respect to their goals to achieve improved health and function based on the criteria they have defined.

Despite the number of studies that have considered the effects of physical activity in seniors, there is evidence to indicate that the full range of effects has not been explored; and there are some data that are missing entirely. Efficacy studies of physical activity that have shown an inverse association with cardiovascular morbidity and mortality (23,, 40, 41, 44), were performed with healthy adults (no underlying cardiovascular disease). Other studies that have attempted to quantify the effects of different components of physical activity (e.g. intensity, frequency, type) have been conducted largely with healthier, younger subjects or have focused on outcomes related to physical fitness (28, 64-67). The applicability of these efficacy studies to

older populations is less well known, particularly for populations with underlying health conditions. A large proportion of seniors do not achieve CDC recommended levels of physical activity, whether or not they are healthy or fit to exercise (1, 68). There is a need to know what levels below those recommended by the Surgeon General would confer benefits that might delay or reverse adverse outcomes (e.g. cognitive impairment, progressive disablement, mortality) in the majority of seniors who appear unable (e.g. pre-existing condition) or unwilling to participate at the recommended levels. Conversely, it is unclear whether the purported benefits of physical activity based on federal guidelines apply to specific key aging outcomes (e.g., cognitive function, physical function) for seniors who achieve the recommended levels of exercise, and not just overall health (35). There is a need to determine if more vigorous levels of activity than those recommended provide additional health-preservation and/or disease-prevention benefits (69). More comprehensive and explicit estimates of the levels of physical activity that are likely to confer benefits across the broad spectrum of functional capacity in older adults are critical, particularly given these individuals adopt fewer (e.g., walking) and more moderate types of LTPA with age, and are less likely to achieve recommended levels of physical activity (1, 68, 70).

Investigators have made attempts to address the multiform ways that physical activity can impact health and well-being in older populations. Studies have examined the effects of physical activity below levels recommended by the CDC, based on types and the number of activities that are more common among older adults (e.g., walking), on important health indicators and outcomes such as body mass index, cholesterol, hypertension, physical function and cognitive status (52-54, 71). Others have examined the joint relationship between physical activity and physical function for subsequent health-related outcomes and found that the associations between physical activity and these outcomes depended on subjects' level of physical function (72). Jerome, *et al.* reported that physical activity was achievable even among those with functional impairments, and prescribed levels of physical activity for such individuals' to avoid progressive disablement (73). Other studies have examined the predictive value gained from knowledge of subjects' current physical function status on their subsequent risk of dementia (55, 74). Ultimately, these studies seek to evaluate the contribution of different causal factors (e.g., physical activity) that underlie the variability of disease outcomes in older populations, and attempt to take into account additional levels of complexity that are associated with the aging process as part of their estimation of these effects. Such studies could shed further light on causal mechanisms of aging and health of older adults, and could provide additional insight with respect to the benefits and potential limitations of the causal factors that occur as part of these mechanisms that might serve as interventions with respect subjects' future health status.

Ideally, one should consider whether the effects of interest on various outcomes have been properly described and to make extensive use of the data that are available. Evaluation of physical activity in the elderly populations, as described previously, is complicated by wide-varying differences in fitness (e.g., cardiovascular) and health of older persons, which are strongly related with levels of physical activity (75). Physical fitness, and changes in fitness and health that may occur in part because of physical activity, can affect seniors' willingness and capacity to engage in future physical activity (76). It is important to separate and determine the relative contributions of these factors (e.g., fitness, physical activity) over the span of the aging process, given that they may have independent effects for different aging outcomes.

Besides physical activity, there are number of factors (e.g., genetic, environmental, social) thought to affect aging outcomes, for which we may want to assess relative effects.

Lifetime exposure to various environmental exposures (e.g., occupational) could have significant implications for aging outcomes, particularly in conjunction with other age-related developments. For example, in addition to active smoking, lifetime exposure to secondhand smoke (SHS) which is not uncommon in the elderly given the widespread practice of cigarette smoking in the U.S. before health effects of tobacco smoke were widely disseminated, has been reported to have major health effects (77). Factors like SHS may indirectly have major health consequences (e.g., dementia) through various health-intermediates known to be related to SHS (e.g., CVD), or be directly linked to particular aging outcomes, like dementia, given the specific risks that they carry (e.g. neurotoxins in cigarette smoke).

Past levels of physical activity and other aspects of the lifecourse, which can influence behavioral- and health-related intermediates (e.g., CVD, subsequent physical activity) in the elderly, may *directly* affect long-term functional outcomes (e.g., cognitive, physical function) and mortality as well. Knowledge of such effects, and their timing, would be informative with respect to particular mechanisms through which physical activity may operate with respect to key aspects of function and aging. Moreover, a developed characterization of the impact physical activity has on aging outcomes should involve an investigation of the modulation of effects of physical activity by underlying fitness and health (e.g., obesity, diabetes, cardiovascular disease). In other words, there may be subpopulations of the elderly for which the effects of physical activity may differ given the known relationships between age-related physiologic changes (e.g., metabolism), health-related intermediates (e.g., diabetes, cardiovascular disease), function (e.g, grip strength, gait speed, cognitive), and mortality (78-82). Accurate causal descriptions of the different factors (e.g., physical activity) that are involved in aging and disablement, and the examination of these factors in the context of other age-related changes (e.g. changes in physical fitness, medication use, onset insulin resistance), is essential for improved understanding that these factors have on long-term health and functioning in the elderly.

Studies that seek a broader context for better understanding the roles of different factors on the aging process, however, may require more sophisticated levels of analysis than conventional analytical methods can provide. For example, investigation of effects of physical activity in the presence of underlying health conditions/disability present particular analytic issues that may lie outside the scope of standard analytical approaches. A review that compiled findings from several observational and experimental studies to examine effects of different physical activity programs specific to different subpopulations of elders with different health histories indicated that most of these studies did not control adequately for factors associated with both disability and physical inactivity and, consequently, could not discern the independence of risk factors and the important causal relationships (75). Intervention studies that have examined the effects of physical activity on subsequent health in seniors with underlying medical conditions have been somewhat informative, but have not addressed the wider scope of factors that would need to be examined (48, 75).

Keysor and Jette (83) reviewed 31 studies that reported on the effects of exercise programs on physical performance as well as disability. While various programs were found to improve strength, aerobic capacity and range of motion, there was little evidence of benefit with respect to disability. The authors noted that a more complex set of factors than had been studied may have deserved consideration in the context of the disablement paradigm. They did not discuss any of the analytical problems related to evaluation of the disablement process (63). In contrast, some epidemiological studies have presented data that support a link between physical activity and disability (43, 84).

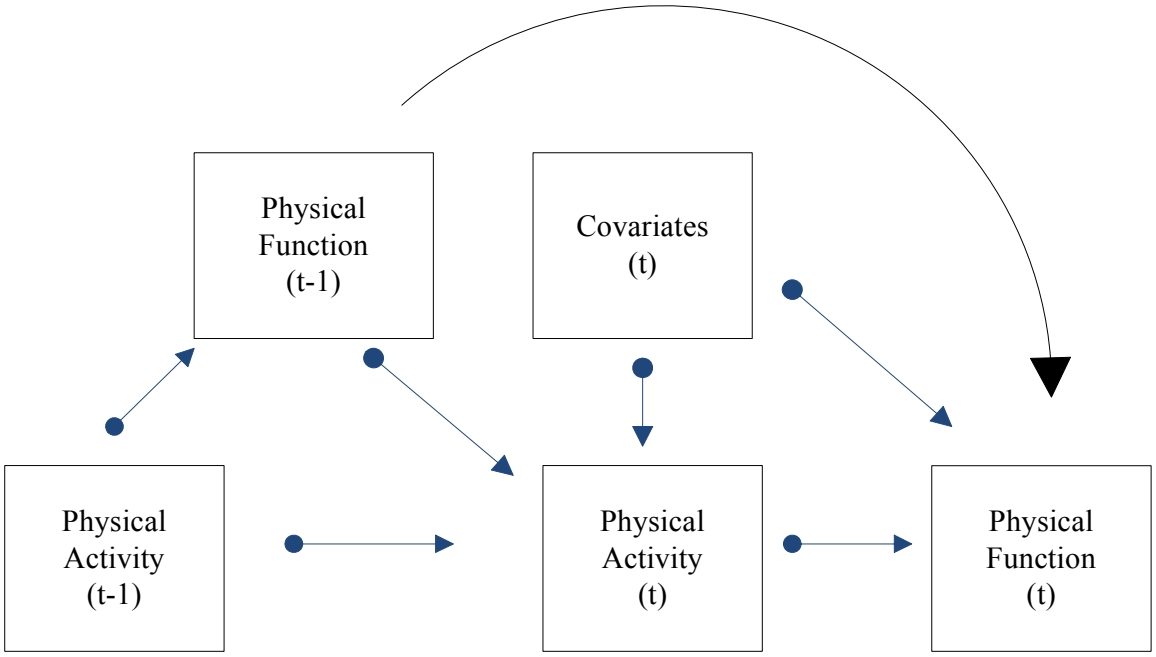
No consensus has been reached with regard to the optimal lifetime exercise habits for maximization of cardiovascular health in the elderly (85) or the effect of exercise in the setting of chronic diseases. The few studies that tried to evaluate the benefits of starting to exercise after a major health event have had mixed results. A review of post-stroke trials reported no association between quality of life and mortality with commencement of physical activities after a stroke (86). Both a cross sectional and a longitudinal study of physical exercise in older persons with chronic heart disease suggested an association (87, 88). These contrasting results indicate clearly that questions remain with regard to the efficacy of exercise begun after the occurrence of different types of cardiovascular events in the elderly.

Based on what has been described from the aforementioned studies, there is no firm consensus with regard to the effects of physical activity and its role for a variety of health outcomes relevant to population aging. In some of the studies discussed, there were contradictory findings with regard to any potential benefit of physical activity, given underlying medical conditions. Keysor and Jette, who described the findings of the studies related to disablement, reasoned that certain factors may not have been considered in the studies that were potentially important (83). Previously, Jette has discussed the influence of different factors related to disablement, and the analytical challenges that confront the researcher with regard to their disentanglement (89). One could conjecture that the exclusion of factors discussed by Keysor and Jette may be related, in part, to the limitations of the analytical techniques that were used in these studies. Singh, in her review of the literature, clearly indicated that many of the studies fell short of accounting for the underlying relationships between factors, and, therefore could not discern properly whether the effects of these were independent (75). Therefore, it may not be for lack of content (i.e., study data) that these studies have fallen short of addressing important concerns related to effects of physical activity on disablement and health, but rather, it may be the inherent limitation of the analytic techniques that have been used for their analysis.

Most observational studies have, at their core, methodological constraints which have not allowed for the proper characterization and quantification of physical activity effects with complex aging cohort data. One example of this is the use of conventional forms of analysis (e.g. logistic models, Cox Proportional hazard models) that have been applied to assess certain effects of interest with respect to function and mortality. These analyses have fallen short with respect to disentanglement of factors—for example, like those described in the disablement process-- due to inability to control for time-dependent confounding (90). Time-dependent confounding refers to the situation in which a factor is a confounder of the exposure and whose level is affected, in turn, by the previous level of the exposure. For example, physical performance (e.g., walking speed) at some time *t* is a risk factor for leisure time physical activity (LTPA, the exposure) at *t* and disability (outcome) at some future time; and the history of LTPA up to *t*, clearly predicts physical performance at *t*. This scenario was addressed by Miller, *et al*. (43) by a series of parallel logistic regression models which, in the end, could not separate fully the effects of physical activity and physical performance limitation on the occurrence of disability. A similar situation occurs for the more complex pathways that involve multiple intermediates and that have the features of time-dependent confounders/intermediates (e.g., pathway from physiological factors through LTPA, medical morbidity, functional performance, mortality). LaCroix, *et al*. (91) used data from the EPESE studies to evaluate factors thought to be related to maintenance of mobility with a proportional hazards model that included physical activity and body mass index (BMI). BMI is both a confounder and on the causal pathway for any effects of physical activity on mobility. Thus, the estimate of the protective effect of physical activity presented

almost certainly is a biased estimate of the marginal (i.e., population) causal influence of physical activity on mobility. More importantly, the regression parameter for LTPA in the Cox model with time-dependent confounders (BMI) has no sensible interpretation—i.e., since the change in level of LTPA implies a change in the level of BMI, one cannot fix the level of either (as is required in such conditional models) to use the model coefficients to estimate an effect of physical activity or BMI on mobility.

FIGURE 2. Depiction of time-dependent confounding in assessment of effects of physical activity on physical function over time. If physical function *(t-1)* is not included in the model, the relationship of physical activity *(t)* with physical function *(t)* will be confounded. However, by adjusting for physical function *(t-1)*, part of the effect of physical activity *(t-1)* on physical function *(t)* is removed.



Causal inference methods account for time-dependent confounding and provide unbiased estimates of effects of interest (includes longitudinal effects) in the presence of other time-varying influences that would otherwise bias estimation of the true effect of interest (90, 92). Application of these methods to observational studies has been shown to quantify correctly known effects where conventional methods provided biased estimates (90, 92-94). Based on these methods, a class of causal statistical models, e.g., marginal structural models (MSMs), may be used to directly model the causal factors of interest. Estimation procedures are applied to under alternative exposure regimens, given the distribution of baseline covariates in a population, which is highly relevant for estimation of likely effects of various interventions in public health/clinical situations. These methods, with respect to #1 and #2 have been applied to the

disablement model and have shown that potential benefits of physical activity on functioning are highly dependent on body composition, with no benefit being observed for marked obesity (95).

A key feature of causal inference methods is that they allow examination of effects in the presence of causal intermediates. Prior to these methods, this issue has represented an important limitation in epidemiologic analyses. For example, to examine an effect that occurs at some point in a process (e.g., cumulative lifetime exposure to secondhand smoke), one has to account for the differences in the outcome due to other factors that could have occurred on the causal pathway (e.g., incident cardiovascular disease) after the exposure of interest. However, by accounting for these factors, one runs the risk of adjusting away the effect of interest. In some situations, it is possible with conventional statistical approaches to obtain unbiased estimates of direct effects that take into account the presence of causal intermediates (96). Alternatively, causal methods allow for the proper disentanglement of intermediates on the causal pathway when conventional approaches fail, and represent a way by which one can assess the independent contributions of the factors involved (96-98).

Randomized controlled trials (RCTs) can be considered as an alternative to observational studies, since these allow for control of factors that can be difficult to account for in the analysis of observational data. Also, RCTs can provide estimates of direct, marginal effects. RCTs have been proposed to develop interventions aimed at preventing and delaying functional decline and disability in older persons (99). However, these studies are costly and limited in their duration to examine long-term effects (100). Also, RCTs may be impossible to conduct for ethical reasons. Moreover, results for intervention studies may be limited in terms of the applicability of their results to individual patients, not only because target populations are not represented typically---particularly in older populations where the exclusion rates would be expected to be high---but also because of the limited number of interventions that can be examined with these studies (101). RCTs typically cannot/do not account for changes with respect to subjects' adherence to treatment regimens (e.g., changes in levels of physical activity), nor can they account for physiologic/behavioral changes that can affect study outcomes and account for heterogeneous responses within the treated and controlled groups (102). Nonetheless, RCTs remain the gold-standard of clinical research to test the effects of interventions. Causal methods as applied to observational studies, if not considered a valid substitute for this gold-standard, can certainly inform the development and planning of RCTs (103).

History-adjusted Marginal Structural Models (HA-MSMs) are a generalization of MSMs that carry all the advantages of traditional MSMs in addition to the ability to model effects of exposures in the presence of changing covariate patterns over time (104, 105). Whereas MSMs can be applied to evaluate the effects of *fixed* treatment regimens (e.g., vigorous vs. moderate exercise) over the whole population (or subpopulations defined at baseline), HAMSMs can be used to define and examine treatment effects *dynamically*, given changes in individuals' covariate (i.e., health) status over time. These treatment effects have greater clinical relevance than MSM treatment effects because the former can be used to determine *optimal* treatment regimens for given underlying health patterns that occur over time. These effects also provide greater insight with respect to patient responsiveness to particular treatment patterns. For example, vigorous levels of physical activity may vary in their effects on cognitive status or mortality in those with and without a new heart attack or stroke. Conversely, the effect of no physical activity on subsequent outcomes could vary by the variable rates of decline in

cardiovascular reserve between subjects. Taken together, these effects can provide clinicians and public health planners with more complete and accurate set of possible outcomes given the levels of physical activity achievable by their patients and populations of elderly, respectively.

Methods that utilize HAMSMs are intuitively easier to comprehend and simpler to implement (e.g. standard software) than other causal approaches that allow for examination of effects in the presence of time-varying covariates (Structural Nested Failure Time Models, Dynamic Treatment Models) (106, 107) and can use observational data to simulate the results of a hypothetical controlled intervention *in the study population that generated the data*. This mitigates one serious limitation of RCTs described above—i.e., those who participate almost always are not representative of their target populations, a situation that results from necessary exclusion criteria, inability to complete the trial by subjects and the fact that most trials are not powered to identify the "responder" sub-sets within the larger group of treated participants- i.e., subgroups of treated subjects and controls are likely to be heterogeneous with respect to their counterfactual outcomes (i.e., outcomes that they would have had if they had been assigned to the opposite group to which they actually were assigned). In addition, most RCTs do not evaluate the effect of time-dependent factors that influence compliance with treatment and or outcomes—i.e., they do not address time-dependent confounding with respect to treatment that occurs after the start of the intervention that might lead to unexpected and/or unmeasured differences between the intervention and control groups.

In summary, most studies to date have not been able to address adequately fundamental questions related to population aging because of limitations of the analytical techniques and/or study designs that have been available to researchers. MSMs and HAMSMs, and causal inference methods in general, represent an advancement over these traditional methods that allow greater flexibility with respect to research questions that can be addressed and the examination of causal effects; the utilization of data available for analysis; and improved control of confounding factors and assessment of response heterogeneity. HAMSMs in particular present an opportunity to determine prescriptive recommendations for physical activity, given changing patterns in health over time, for preservation of long-term physical and cognitive function. In addition, though these methods may not be considered a substitute for RCTs, given these methods are implemented with observational data, they nonetheless may be highly informative in the development of RCTs. These methods are described and applied to three important questions of population aging which would have been impossible to address with conventional analyses.

References

1. NIA. 65+ in the United States: 2005: U.S. Dept of Health and Human Services (NIH/NIA)---P23-209, 2005.
2. Congressional Research Service. Life Expectancy in the United States: Library of Congress, 2006.
3. Rowe JW, Kahn RL. Successful Aging. New York: Pantheon Books, 1998.
4. Brooks G. Exercise Physiology: McGraw-Hill, 2005.
5. Department of Health and Human Services. 7th Report Joint National Commitee on Prevention Detection Evalutation and Treatment of High Blood Pressure, 2004.

6. Bortz WM, 2nd. A conceptual framework of frailty: a review. J Gerontol A Biol Sci Med Sci 2002;57:M283-8.
7. Aalami OO, Fang TD, Song HM, Nacamuli RP. Physiological features of aging persons. Arch Surg 2003;138:1068-76.
8. Hayflick L. Biological Aging Is No Longer An Unsolved Problem. Annals of New York Academy of Science 2007;1100:1-13.
9. Rogers MA, Hagberg JM, Martin WH, 3rd, Ehsani AA, Holloszy JO. Decline in VO2max with aging in master athletes and sedentary men. J Appl Physiol 1990;68:2195-9.
10. Fleg JL, Morrell CH, Bos AG, et al. Accelerated longitudinal decline of aerobic capacity in healthy older adults. Circulation 2005;112:674-82.
11. Hollenberg M, Yang J, Haight TJ, Tager IB. Longitudinal changes in aerobic capacity: implications for concepts of aging. J Gerontol A Biol Sci Med Sci 2006;61:851-8.
12. Larsson L. The age-related motor disability: underlying mechanisms in skeletal muscle at the motor unit, cellular and molecular level. Acta Physiol Scand 1998;163:s27-s29.
13. Roubenoff R, Hughes VA. Sarcopenia: current concepts. Journal of Gerontology A: Biological Science and Medical Science 2000;55:M716-24.
14. Williams GN, Higgins MJ, Lewek MD. Aging skeletal muscle: physiologic changes and the effects of training. Phys Ther 2002;82:62-8.
15. Vandervoort A. Aging of the Human Neuromuscular System. Muscle Nerve 2002;25:17-25.
16. Goldspink DF. Ageing and activity: their effects on the functional reserve capacities of the heart and vascular smooth and skeletal muscles. Ergonomics 2005;48:1334-51.
17. Tager IB, Hollenberg M, Satariano WA. Association between self-reported leisure-time physical activity and measures of cardiorespiratory fitness in an elderly population. American Journal of Epidemiology 1998;147:921-31.
18. Drewnowski A, Evans WJ. Nutrition, physical activity, and quality of life in older adults: summary. J Gerontol A Biol Sci Med Sci 2001;56 Spec No 2:89-94.
19. Tager IB, Haight TJ, Hollenberg M, Satariano WA. Physical functioning and mortality in older women: Assessment of energy expenditure and level of difficulty. Journal of Clinical Epidemiology 2003;56:807-813.
20. Weinert BT, Timiras PS. Invited review: Theories of aging. J Appl Physiol 2003;95:1706-16.
21. Kramer AF, Erickson KI, Colcombe SJ. Exercise, cognition, and the aging brain. J Appl Physiol 2006;101:1237-42.
22. Mattson MP, Magnus T. Ageing and neuronal vulnerability. Nat Rev Neurosci 2006;7:278-94.
23. Centers for Disease Control. Physical Activity and Health: A Report of the Surgeon General. Pittsburgh: Dept. of Health and Human Services--S/N017-123-00196-5, 1996.
24. Haskell WL, Lee IM, Pate RR, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Circulation 2007;116:1081-93.
25. Bauman AE. Updating the evidence that physical activity is good for health: an epidemiological review 2000-2003. J Sci Med Sport 2004;7:6-19.
26. Lee IM. Dose-response relation between physical activity and fitness: even a little is good; more is better. Jama 2007;297:2137-9.

27. Thompson PD. Exercise and physical activity in the prevention and treatment of atherosclerotic cardiovascular disease. Arterioscler Thromb Vasc Biol 2003;23:1319-21.
28. Franco OH, de Laet C, Peeters A, Jonker J, Mackenbach J, Nusselder W. Effects of Physical Activity on Life Expectancy With Cardiovascular Disease. Arch Intern Med 2005;165:2355-2360.
29. Seeman T, Rodin J, Albert M. Self-Efficacy and Cognitive Performance in High-Functioning Older Individuals: MacArthur Studies of Successful Aging. Journal of Aging and Health 1993;5:455-474.
30. Wang C, Satariano W. Self-Rated Current and Future Health Independently Predict Subsequent Mortality in an Aging Population. Journal of Gerontology 2007;62A:1428-1434.
31. Cotman C, Berchtold N. Exercise: a behavioral intervention to enhance brain health and plasticity. Trends Neuroscience 2002:295-301.
32. Cotman C, Berchtold N. Physical activity and the maintenance of cognition: Learning from animal models. Alzheimer's & Dementia 2007;3:S30-S37.
33. Yankner BA, Lu T, Loerch P. The aging brain. Annu Rev Pathol 2008;3:41-66.
34. Mazzeo R, Cavanagh P, Evans WJ, et al. Exercise and Physical Activity for Older Adults. Medicine and Science in Sports and Exercise 1998;30.
35. Nelson ME, Rejeski WJ, Blair SN, et al. Physical Activity and Public Health in Older Adults: Recommendation From the American College of Sports Medicine and the American Heart Association. Circulation 2007;116:1094-1105.
36. Kaplan GA, Seeman TE, Cohen RD, Knudsen LP, Guralnik J. Mortality among the elderly in the Alameda County Study: Behavioral and demographic risk factors. American Journal of Public Health 1987;77:307-12.
37. Berkman LF, Seeman TE, Albert M, et al. High, usual and impaired functioning in community-dwelling older men and women: findings from the MacArthur Foundation Research Network on Successful Aging. Journal of Clinical Epidemiology 1993;46:1129-40.
38. Seeman TE, Berkman LF, Charpentier PA, Blazer DG, Albert MA, Tinetti ME. Behavioral and psychosocial predictors of physical performance: MacArthur Studies of Successful Aging. Journal of Gerontology: Medical Sciences 1995;50A:M177-83.
39. DiPietro L. The epidemiology of physical activity and physical function in older people. Medicine and Science in Sports and Exercise 1996;28:596-600.
40. LaCroix AZ, Leveille SG, Hecht JA, Crothaus LC, Wagner EH. Does walking decrease the risk of cardiovascular disease hospitalizations and death in older adults? Journal of the American Geriatric Society 1996;44:113-20.
41. Kushi LH, Fee RM, Folsom AR, Mink PJ, Anderson KE, Sellers TA. Physical activity and mortality in postmenopausal women. Journal of the American Medical Association 1997;277:1287-92.
42. Fried LP, Kronmal RA, Newman AB, et al. Risk factors for 5-year mortality in older adults. The Cardiovascular Health Study. Journal of the American Medical Association 1998;279:585-592.
43. Miller ME, Rejeski WJ, Beboussin BA, Ten Have TR, Ettinger WH. Physical activity, functional limitation, and disability in older adults. Journal of the American Geriatrics Society 2000;48:1264-72.

44. Rockhill B, Willett WC, Manson JE, et al. Physical activity and mortality: A prospective study among women. Americal Journal of Public Health 2001;91:578-83.
45. Ekelund U, Brage S, Franks PW, et al. Physical activity energy expenditure predicts changes in body composition in middle-aged healthy whites: effect modification by age. Am J Clin Nutr 2005;81:964-969.
46. Stewart KJ, Bacher AC, Turner KL, et al. Effect of Exercise on Blood Pressure in Older Persons: A Randomized Controlled Trial. Arch Intern Med 2005;165:756-762.
47. Janssen I, Jolliffe CJ. Influence of physical activity on mortality in elderly with coronary artery disease. Med Sci Sports Exerc 2006;38:418-7.
48. Gregg EW, Brown A. Cognitive and Physical Disabilities and Aging-Related Complications of Diabetes. Clinical Diabetes 2003;21:113-118.
49. Yaffe K, Barnes D, Nevitt M, Lui L-Y, Covinsky K. A prospective study of physical activity and cognitive decline in elderly woman. Archives of Internal Medicine 2001;161:1703-08.
50. Laurin D, Verreault R, Lindsay J, MacPherson K, Rockwood K. Physical activity and risk of cognitive impairment and dementia in elderly persons. Arch Neurol 2001;58:498-504.
51. Barnes DE, Yaffe K, Satariano WA, Tager IB. A Longitudinal Study of Cardiorespiratory Fitness and Cognitive Function in Healthy Older Adults. Journal of the American Geriatrics Society 2003;51:459-465.
52. Abbott RD, White LR, Ross GW, Masaki KH, Curb JD, Petrovitch H. Walking and dementia in physically capable elderly men. Jama 2004;292:1447-53.
53. Weuve J, Kang JH, Manson JE, Breteler MM, Ware JH, Grodstein F. Physical activity, including walking, and cognitive function in older women. Jama 2004;292:1454-61.
54. Podewils LJ, Guallar E, Kuller LH, et al. Physical activity, APOE genotype, and dementia risk: findings from the Cardiovascular Health Cognition Study. Am J Epidemiol 2005;161:639-51.
55. Wang L, Larson EB, Bowen JD, van Belle G. Performance-based physical function and future dementia in older people. Arch Intern Med 2006;166:1115-20.
56. Almeida OP, Norman P, Hankey G, Jamrozik K, Flicker L. Successful Mental Health Aging: Results From a Longitudinal Study of Older Australian Men. Am. J. Geriatr. Psychiatry 2006;14:27-35.
57. Jagust W, Harvey D, Mungas D, Haan M. Central obesity and the aging brain. Arch Neurol 2005;62:1545-8.
58. McEwen BS. Physiology and Neurobiology of Stress and Adaptation: Central Role of the Brain. Physiologic Reviews 2007;87:873-904.
59. Wu W, Brickman AM, Luchsinger J, et al. The brain in the age of old: the hippocampal formation is targeted differentially by diseases of late life. Ann Neurol 2008;64:698-706.
60. Colcombe SJ, Erickson KI, Raz N, et al. Aerobic Fitness Reduces Brain Tissue Loss in Aging Humans. J Gerontol A Biol Sci Med Sci 2003;58:M176-180.
61. Spence AP. Biology of Human Aging: Prentice Hall, 1999.
62. Sayer AA, Cooper C. Aging, sarcopenia and the life-course. Reviews in Clinical Gerontology 2006;16:265-274.
63. Verbrugge LM, Jette AM. The disablement process. Social Science and Medicine 1994;38:1-14.

64. Manson JE, Hu FB, Rich-Edwards JW, et al. A prospective study of walking as compared with vigorous exercise in the prevention of coronary heart disease in women. N Engl J Med 1999;341:650-8.
65. Sesso HD, Paffenbarger RS, Jr, Lee I-M. Physical Activity and Coronary Heart Disease in Men : The Harvard Alumni Health Study. Circulation 2000;102:975-980.
66. Lee IM, Sesso HD, Oguma Y, Paffenbarger RS, Jr. The "weekend warrior" and risk of mortality. Am J Epidemiol 2004;160:636-41.
67. Duncan GE, Anton SD, Sydeman SJ, et al. Prescribing exercise at varied levels of intensity and frequency: a randomized trial. Arch Intern Med 2005;165:2362-9.
68. CDC. Leisure-Time Physical Activity Among Adults: United States 1997-98. Advance Data from Vital and Health Statistics 2002.
69. Williams PT. Relationships of heart disease risk factors to exercise quantity and intensity. Archives of Internal Medicine 1998;158:237-45.
70. DiPietro L. Physical Activity in Aging: Changes in Patterns and Their Relationship to Health and Function. J Gerontol A Biol Sci Med Sci 2001;56:13-22.
71. Mensink G, Ziese T, Kok F. Benefits of leisure-time physical activity on the cardiovascular risk profile at older age. International Journal of Epidemiology 1999;28:659-666.
72. Taaffe DR, Henwood TR, Nalls MA, Walker DG, Lang TF, Harris TB. Alterations in Muscle Attenuation following Detraining and Retraining in Resistance-Trained Older Adults. Gerontology 2008.
73. Jerome GJ, Glass TA, Mielke M, Xue QL, Andersen RE, Fried LP. Physical activity participation by presence and type of functional deficits in older women: The Women's Health and Aging Studies. J Gerontol A Biol Sci Med Sci 2006;61:1171-6.
74. Larson EB, Wang L, Bowen JD, et al. Exercise is associated with reduced risk for incident dementia among persons 65 years of age and older. Ann Intern Med 2006;144:73-81.
75. Singh MA. Exercise comes of age: rationale and recommendations for a geriatric exercise prescription. J Gerontol A Biol Sci Med Sci 2002;57:M262-82.
76. Satariano WA, Haight TJ, Tager IB. Reasons given by older people for limitation or avoidance of leisure time physical activity. J Am Geriatr Soc 2000;48:505-12.
77. Llewellyn DJ, Lang IA, Langa KM, Naughton F, Matthews FE. Exposure to secondhand smoke and cognitive impairment in non-smokers: national cross sectional study with cotinine measurement. Bmj 2009;338:b462.
78. Luchsinger JA, Reitz C, Patel B, Tang MX, Manly JJ, Mayeux R. Relation of diabetes to mild cognitive impairment. Arch Neurol 2007;64:570-5.
79. Rosano C, Brach J, Studenski S, Longstreth WT, Jr., Newman AB. Gait variability is associated with subclinical brain vascular abnormalities in high-functioning older adults. Neuroepidemiology 2007;29:193-200.
80. Sayer AA, Syddall HE, Dennison EM, et al. Grip strength and the metabolic syndrome: findings from the Hertfordshire Cohort Study. Qjm 2007;100:707-13.
81. Gale CR, Martyn CN, Cooper C, Sayer AA. Grip strength, body composition, and mortality. Int. J. Epidemiol. 2007;36:228-235.
82. Wright CB, Festa JR, Paik MC, et al. White matter hyperintensities and subclinical infarction: associations with psychomotor speed and cognitive flexibility. Stroke 2008;39:800-5.

83. Keysor JJ, Jette AM. Have we oversold the benefit of late-life exercise? Journal of Gerontology: Medical Sciences 2001;7:M412-M423.
84. Simonsick EM, Lafferty ME, Phillips CL, et al. Risk due to inactivity in physically capable older adults. American Journal of Public Health 1993;83:1443-50.
85. Twisk JW, Kemper HC, van Mechelen W. Prediction of cardiovascular disease risk factors later in life by physical activity and physical fitness in youth: introduction. Int J Sports Med 2002;23 Suppl 1:S5-7.
86. Meek C, Pollock A, Potter J, Langhorne P. A Systematic Review of exercise trials post stroke. Clin Rehab 2003;17:6-13.
87. Oldridge NB, Guyatt GH, Fischer ME, Rimm AA. Cardiac rehabilitation after myocardial infarction. Combined experience of randomized clinical trials. Jama 1988;260:945-50.
88. O'Connor G, Buring J, Yusuf S, et al. An overview of randomized trials of rehabilitation with exercise after myocardial infarction. Circulation 1989;80:234-244.
89. Jette AM. Disentagling the process of disablement. Social Science in Medicine 1997;48:471-2.
90. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550-60.
91. LaCroix AZ, Guralnik JM, Berkman LF, Wallace RB, Satterfield S. Maintaining mobility in later life II. smoking, alcohol consumption, physical activity, and body mass index. American Journal of Epidemiology 1993;137:858-869.
92. Hernan M, Brumback B, Robins J. Marginal Structural Models to Estimate Causal Effect of Zidovudine on the Survival of HIV-Positive Men. Epidemiology 2000;11:561-570.
93. Cole SR, Hernan MA, Margolick JB, Cohen MH, Robins JM. Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on CD4 cell count. Am J Epidemiol 2005;162:471-8.
94. Mortimer KM, Neugebauer R, van der Laan M, Tager IB. An application of model-fitting procedures for marginal structural models. Am J Epidemiol 2005;162:382-8.
95. Tager IB, Haight TJ, Yu Z, Sternfeld B, van der Laan MJ. Longitudinal study of the effects of physical activity and body composition on functional limitation in the elderly. Epidemiology 2004;15:479-493.
96. Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. Epidemiology 2006;17:276-84.
97. Eisner M, Wang Y, Haight T, Balmes J, Hammond K, Tager I. Second-hand smoke exposure, pulmonary function, and cardiovascular mortality. Annals of Epidemiology 2007;17:364-373.
98. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. Epidemiology 2009;20:18-26.
99. Ferrucci L, Guralnik JM, Studenski S, Fried LP, Cutler GB, Jr., Walston JD. Designing randomized, controlled trials aimed at preventing or delaying functional decline and disability in frail, older persons: a consensus report. J Am Geriatr Soc 2004;52:625-34.
100. Pincus T. Limitations of randomized clinical trials to recognize possible advantages of combination therapies in rheumatic diseases. Semin Arthritis Rheum 1993;23:2-10.
101. Collet JP. [Limitations of clinical trials]. Rev Prat 2000;50:833-7.
102. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. Jama 2007;298:1209-12.

103. Petersen M, Deeks S, van der Laan M. Individualized Treatment Rules: Generating Candidate Clinical Trials. Statistics in Medicine 2007;26:4578-4601.
104. van der Laan M, Petersen M, Joffe M. History-Adjusted Marginal Structural Models and Statically-Optimal Dynamic Treatment Regimens. International Journal of Biostatistics 2005;1.
105. Petersen M, Deeks S, Martin J, van der Laan MJ. History-adjusted Marginal Structural Models for Estimating Time-varying Effect Modification. American Journal of Epidemiology 2006;166:985-993.
106. Robins JM. Marginal Structural Models versus Structural Nested Models as Tools for Causal Inference. In: Halloran ME, Berry D, eds. Statistical Models in Epidemiology: The Environment and Clinical Trials. New York: Springer-Verlag, 1999:95-134.
107. Murphy SA. Optimal dynamic treatment regimes. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2003;65:331-355.

Chapter 2 —Study Methodology (General)

The previous chapter described the rationale (the 'why') for the application of causal inference methods (aka 'causal statistical methods') in epidemiologic aging research, and their potential to advance the field quantitatively. In this chapter, I describe the 'how' of causal statistical methods: first, the means by which these methods are used to identify and quantify causal effects based on observational data; and second, with regard to their applicability to address a wide-range of causal questions of aging.

The purpose of epidemiologic research and data analysis is concerned with the examination of causal questions related to potential risks factors (causes) and disease outcomes (effects). Traditionally, in non-experimental (observational) studies, analyses are carried out which examine the association of risk factors and disease incidence, controlling for confounders and other potential sources of bias (e.g., selection bias, measurement error). While these types of analyses are useful, they represent descriptive analyses of the joint distributions of the variables of interest. They do not address the underlying causal question of interest posited by the epidemiologist, which is: "For a given change in level of risk factor *x* in the population, what is the change in outcome status *y*?" Such questions are not testable directly, based on observational data, because one does not observe more than one level of a risk factor, and one outcome that corresponds with this risk factor, in any given individual at a given time (1).

Marginal Structural Models (MSMs) represent one class of causal models that can be utilized for directly defining and assessing causal effects of interest with observational data. These models are based on the notion of counterfactuals, which represents the set of outcomes subjects might have experienced if they experienced exposures (treatments) other than the ones actually received (2, 3). Hypothetically, if one knew the outcomes that corresponded to all possible exposures that a given subject could experience (i.e., the outcomes associated with their 'actual' exposure experience as well as their 'counterfactual' exposure experiences), then, each subject would serve as his/her own control and one could assess whether differences in the outcome were attributable causally to differences in the level of exposure. In practice, one does not observe all possible outcomes. MSMs are examined as part of an analytical framework comprised of assumptions and estimation procedures to recreate the conditions of the ideal study described above where observed data can be evaluated as counterfactual data.

Part of this analytical framework is based on principles of exchangeability and experimentation, both of which are required to identify causal effects. With standard analytical methods, exchangeability is typically evoked by the control of a set of cofactors that might otherwise distort the association of interest. However, it is unlikely that this principle is satisfied typically and presents important limitations for conventional analyses, given that: 1) distributions of cofactors that are included at a study's outset are likely to change over time; and 2) the usual analytical methods used to evaluate effects evaluate change in one variable occurs while all other covariates in the model remain constant, and, therefore, cannot address time-dependent confounding and the complexity of time-dependent interrelatedness of variables in longitudinal data analyses. Moreover, conventional analyses typically ignore the issue of experimentation. Given that several variables may be controlled simultaneously in such traditional analyses, it is unlikely that all of the different levels of the risk factor under

examination are represented in the different strata formed by these cofactors; therefore, the ability to detect differences in outcome with respect to levels of risk is limited. By contrast, causal statistical methods address issues of exchangeability and experimentation, which not only serve to identify causal effects, but provide a more rigorous systematic approach for analyses over conventional methods. The following sections describe in greater detail the evocation of these principles within the analytical framework of assumptions and estimation procedures that underlie causal statistical methods.

Assumptions

I begin with a discussion of how MSM analyses are used to recreate the ideal study—of evaluating observational data as though counterfactual data were available for each subject. Observational data are characterized typically by one treatment and one outcome per subject, and a set of covariates—or, if the data are longitudinal, as one set of covariates, a treatment and correspondent outcome measured at each time point for the length of the study. One of the assumptions that underlies the use of MSMs is the *potential outcomes assumption*. This assumption stipulates that counterfactual or potential outcomes data exist for each and every subject, even though such outcomes are not actually observed (4). In connection with this assumption is the *consistency assumption*, which states that each subject's observed data represent one realization of his/her counterfactual data. For example, in a point treatment (i.e., single time point) study, we define the observed data, *O=(W, A, Y=Y(A)),* where *"W"* represents the observed baseline covariates, "*A*", the treatment (exposure) assignment, and *Y(A)*, the outcome under observed treatment "*A*". The observed data *O=(W,A,Y)* on a randomly sampled subject represent one realization/component of the counterfactual "full" data $X=((Y(a),a\in \mathsf{A}),W)$ for that individual when exposure $a=A$. Thus, the observed data can be characterized as a full data structure *X* with missing data to represent the counterfactual data that are never actually observed(3).

The extension of the framework to longitudinal data structures is straightforward. For each subject in a longitudinal study, followed over time *j=0,…,k,* where $k^*\leq k$ denotes right censoring time, one observes: $L(0),A(0),Y(0),...L(k^*),A(k^*),Y(k^*)$, where $A(j)$ is "treatment" at time *j*; $L(j)$ denotes all time-dependent confounders, and $Y(j)$ denotes the outcome process of interest. In this framework, it is important to note that *Y(j)* is considered a covariate—i.e., *Y(j)* can influence the treatment and future occurrences of Y. *It is assumed that the variables listed above are time-ordered—i.e., A(j),L(j) precede* $Y(k^*)$. Given a treatment regimen $a=(a(0),\ldots,a(k))$, the data can be redefined in terms of counterfactuals, $X_a(j) = (L_a(j), Y_a(j))$, as the data process that would have been observed if the subject, possibly contrary to fact, would have had "treatment" regimen $\boldsymbol{a}$ (e.g. high-level LTPA rather than no LTPA). Here, the data at time *j*, $X_a(j)$ (the subscript '*a*' is used to refer to a counterfactual), only depend on the past treatment, *a(1),…,a(j)*. In this context, the observed data correspond to observation of the "treatment"-specific process ($A(0), X_A(0),\ldots A(k^*), X_A(k^*)$) which corresponds with the <u>actual</u> treatment *A* experienced by the subject and was observed in the study—which *is only one realization* of the data processes that could have been experienced by the subject (3).

The fact that a subject's data are comprised of one line of observed data with the remainder of his/her counterfactual data missing constitutes a missing data problem, if one is interested in recreating conditions of "full" counterfactual data. The problem is addressed partly through the potential outcomes assumption and consistency assumption. However, in addition,

the treatment or exposure that a subject receives at a given time *j* is not at random. In other words, the outcome that one observes may not be the effect of the treatment alone, but may also be, and frequently is in observational data, the result of extraneous factors (e.g., medical events) both measured and unmeasured in the data. Therefore, one of the key assumptions that underlies the identification of causal effects with observational data is based on the principle of exchangeability—i.e., that the treatment process is independent, or randomized, with respect to the outcome process at any given time, $Y_{\bar{a}}(j) \perp \bar{A}(j)$ (2, 3). This assumption, aptly named the *sequential randomization assumption* (SRA), states that 1) there is no confounding by unmeasured covariates, and 2) among the measured covariates, at each time point, treatment is randomized with respect to the outcome conditional on covariate history and past treatment history up to time j : $Y_{\bar{a}}(j) \perp A(j) \mid \bar{L}(j), \bar{A}(j-1)$ (2, 3).

Another way of thinking about this assumption is that, within strata defined by covariate and treatment history $\left(\bar{L}(j), \bar{A}(j-1)\right)$, *A(j)*, or the treatment the subject receives at *j*, is randomized with respect to $Y_a(j)$. This randomization can be depicted graphically in the following DAG (Figure 1).

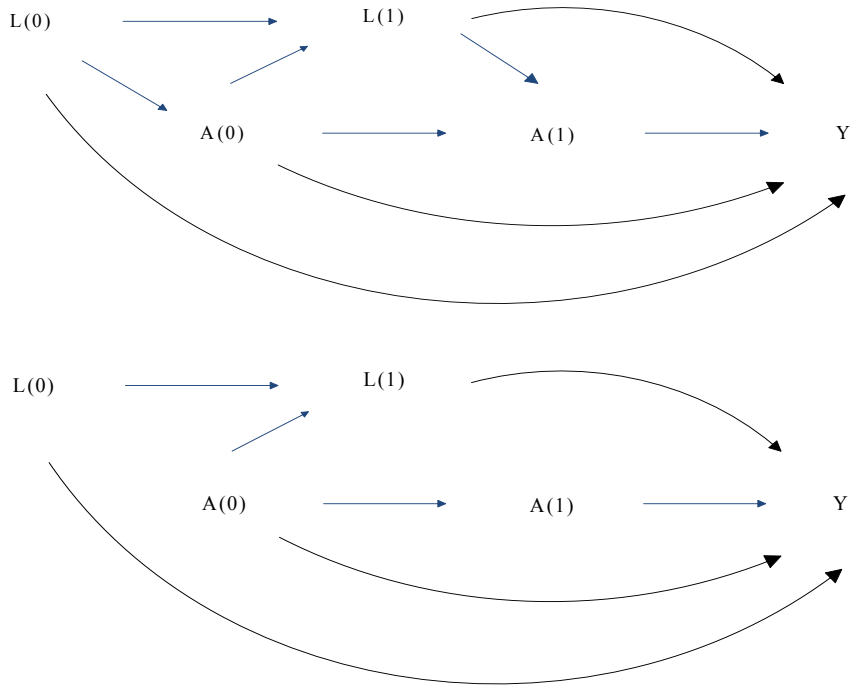


Figure 1. Representation of time-dependent confounding of treatment process *A(j)* with respect to outcome Y for times *j=0,1*. In the top graph, the arrows from *L(j)* to *A(j) and Y* represent confounding of *A* and *Y* by time-dependent covariates (the graph assumes no unmeasured confounders). *Y* is assumed to single endpoint in time.The bottom graph shows that by removal of the arrows from *L(j)* to *A(j)*, the treatment process *A(j)* is "randomized" with respect to the outcome.

In order for randomization to apply over all time points---e.g., if one is interested in the causal effect of a series of treatments or "treatment history", rather than the effect of a single treatment---one can restate the SRA for all timepoints j as $g(A \mid X) = \prod_{j=0}^{k} g(A(j) \mid \bar{A}(j-1), \bar{L}(j))$, where the left side of the equation $g(A \mid X)$ expresses randomization of A with respect to the full data X *as being the equivalent to* the product of a function which relates conditional probabilities of $A(j)$ given past treatment $\bar{A}(j-1)$ and covariate history $\bar{L}(j)$ at each time j. Proper specification of this function $g(A \mid X)$ is one approach to satisfying the SRA---i.e., beyond past treatment and the measured covariates, the relationship of A is random with respect to Y, and is achieved through one of the estimation procedures, described in the next section, to estimate MSM parameters.

Finally, among the different assumptions required to identify causal effects is the experimental treatment assignment (ETA), or positivity, assumption(3). This assumption, based on the principle of experimentation, states that, for all given covariate patterns in the data, all treatment levels are possible (i.e., observed). This assumption can be expressed as:

$$\Pr(A(j) = a \mid \bar{A}(j-1), \bar{L}(j), \bar{Y}(j-1)) > 0 \text{ for all } a \in \{a(j) : a \in \mathrm{A}, \bar{a}(j-1) = \bar{A}(j-1)\}$$

In other words, all subjects (given their covariate levels) have a positive probability of assignment to the different treatment levels under consideration. To the extent that the assumption is not satisfied (e.g., no high aerobic activity among unhealthy subjects), one can not assess potential causal effects of treatment, because certain levels of treatment simply do not occur, or occur with very low probability—i.e., there is no experimentation at that level of treatment. To some degree, all the estimation procedures used to quantify causal effects (described below) depend on satisfaction of the ETA assumption. For example, G-computation will still provide consistent estimates, but ones which are less efficient (i.e., increased variance), if the assumption is violated. IPTW estimation, in particular, is susceptible to bias given ETA violation. For other estimators with double-robust properties that model different parts of the likelihood of the data (see below) to estimate causal parameters, ETA violation signifies that the *g*-part of the likelihood is not identifiable, so that correct specification of the other "identifiable" portion of likelihood (*Fx*-part) is required to obtain consistent MSM estimates.

Methods are available to examine both the extent of ETA violation and to address the reduced efficiency and potential bias of MSM parameters, if, indeed, such a violation occurs. Details are given elsewhere(5, 6).

Estimation

Overview

Under the potential outcomes assumption, the consistency assumption, the temporal ordering assumption, the SRA assumption, and the ETA assumption, as described above, causal parameters are identifiable with both full and observed data. The likelihood of observed data can be formulated as below. Here, *A(j)* represents a time-varying treatment, *L(j)* represents time-varying covariates, which also includes the outcome measure *Y(j)*, which itself can be thought of as a time-varying covariate at *jth* time in the process-i.e., *Y(j)= L(j+1)*. Implicit in this likelihood

are causal effects of interest that relate *Y(j)* and *A(j)*, but which are confounded by other elements, namely *L(j)*. However, the SRA implies factorization of the observed data likelihood *L(O)* into two parts, shown below, referred to as the *g* and the *FX* parts(7):

$$L(O) = f(L(0)) \underbrace{\prod_{j=1}^{K+1} f(L(j) \mid \bar{L}(j-1), \bar{A}(j-1))}_{FX \text{ part}} \underbrace{g(\bar{A}(j) \mid X)}_{g \text{ part}}$$

Consistent estimation of either the *FX* part of the likelihood using G-computation (G-Comp), or the *g*-part of the likelihood using inverse probability of treatment (IPTW) are two approaches for satisfaction of the SRA(2, 7, 8). A double-robust IPTW estimator (DR-IPTW), which relies on the specification of the *Fx* part and the *g*-part of the likelihood, but only requires proper specification of one of these parts, is also available(9). An alternative estimator —known as the Targeted Maximum Likelihood Estimator (TMLE), has double-robust features, given that it relies on separate models of the *FX* part and the *g*-part of the likelihood(10). However, compared to DR-IPTW, it is a simpler estimator to implement, while maximizing the gains to be obtained from being doubly-robust(10, 11). Each of these estimators is discussed below in greater detail.

Recently, data-adaptive model selection algorithms have been developed to quantify more accurately different parts of the likelihood and satisfaction of the SRA, which is required for consistent estimates of causal parameters(12-14). A detailed overview and application of one type of data-adaptive procedure (cross-validation Deletion/Substitution/Addition, or cross-validation DSA algorithm) for model selection for: 1) satisfaction of the SRA and 2) the MSM itself—i.e., models of causal exposure-response relationships given available data distributions, is provided in the Appendix 1.

G-computation Estimation

One approach to satisfaction of the SRA is to build a model where *Y* is regressed on *A* and all measurable confounders of *Y* and *A*—the so-called *FX* part of the likelihood, such that the effects of *A* are randomized with respect to Y within the controlled factors(3, 8). This approach underlies traditional multivariable regression, where the effect of interest is adjusted by a set of controlling factors in the regression model. Although the G-Computation estimator can be applied to estimate effects with longitudinal data, one can demonstrate its implementation in a point treatment setting. The extension to longitudinal data is straightforward. For example, the regression can be represented by $E(Y \mid A = a, W = w)$ where W represents all measured confounders of *Y* and *A*. If the assumption of no unmeasured confounding can be assumed reasonably to be true, and the model is properly specified (e.g., application of cross-validation DSA that does not depend on assumptions with regard to model form), then $E(Y \mid A = a, W = w) = E(Y_a \mid W = w)$ so that the parameter in front of *A* in the model represents an adjusted causal effect. Given that the parameter is adjusted, or stratum-specific---i.e., effect of *A* within strata defined by *W*---- we may choose to obtain an overall causal effect of *A* which can be done by summing, or marginalizing, the effect of *A* over all strata *W*=*w*: $E(Y_a) =$

$\int_w E(Y \mid A,W=w)df(W=w)$, if W is continuous, or $\sum_w E(Y \mid A,W=w)P(W=w)$ if W is discrete. Alternatively, one may choose to examine stratum-specific effects of A for a subset of variables V of W ($V \subset W$), by marginalizing the effect of A over strata W that do not include V.

IPTW Estimation

By comparison with G-computation, one can specify instead a model for the *g*-portion of the likelihood described above, the so-called *treatment mechanism*. To satisfy the SRA, a correct specification of treatment assignment at time *j*, given past treatment and covariate history (covariates that affect treatment assignment at a given time *j*), is necessary. For example, the treatment mechanism $g(A \mid X)$, rewritten as $\prod_{j=1}^{K+1}(g(\overline{A}(j) \mid X)$ given j time points, can be expressed as $\prod_{j=0}^{k} g(A(j) \mid \overline{A}(j-1),\overline{L}(j))(2)$. This is no more than the product of the conditional probabilities of *A(j)* given the past for each time *j*. Therefore, for each *j*, a conditional model of *A(j)* given the past, with data from all subjects, can be developed. There are a variety of tools for fitting the treatment mechanism, but the underlying assumption is that the model is properly specified. Again, a model selection tool, like cross-validation DSA, is preferred since is not dependent on assumptions about the form of the model used for the treatment model.

At the next step, the fitted treatment mechanism is used as part of an estimating equation to estimate the causal model(2, 3). In practice, this is equivalent to a weighted regression of the causal parameter of interest and baseline covariates V ($V \subset L(0)$) with the observed data, where each subject contributes a weight for each time point she/he has observed values in the data. These weights, in general terms, represent the inverse probability of a subject being assigned the set of treatments, or treatment regimen $\overline{A}$, until time *j* given past treatment and covariate history expressed as:

$$SW_{ij} = \frac{\prod_{j=0}^{k} g(A_i(j) \mid \overline{A_i}(j-1),V)}{\prod_{j=0}^{k} g(A_i(j) \mid \overline{A}_i(j-1),\overline{L_i}(j))} \quad i=1,\ldots,n,\ j=0,\ldots,k$$

An interpretation of these weights is as follows: the probability that an individual receives treatment A given his/her previous treatment and covariates is a reflection of the level of confounding of *A* and *outcome Y* by the past---i.e., probability of high LTPA and a favorable outcome is greater among those in good health. Thus, the weight is a numeric representation of the level of confounding of *A* by past variables. By inclusion of the weights in the regression, one accounts for the influence of the past with respect to *A* (see Figure 3 above). For example, healthy individuals who experience high LTPA are overrepresented in the data, and so these individuals' probability of high LTPA is large. Conversely, unhealthy individuals who have low or no LTPA are likely to be overrepresented as well. On the other hand, there may be individuals who might actually be observed to have high LTPA, for example, but based on their covariate

patterns would be expected to have a lower probability of high physical activity (e.g. older, unhealthy subjects). These individuals, in contrast to the former, would receive greater weights in the data (by inverse probability), and would offset the individuals who are overrepresented. Consequently, any association of health and LTPA that existed previously in the observed data is removed. If *A* is unconfounded in the original data initially, the weights would equal 1, and the regression of the causal parameter described above could be carried out because *A* would be "randomized" with respect to *Y*.

Intuitively, inclusion of *V* in the weight expression above makes sense. By the inclusion of *V* in the causal model itself, one can partly account for the confounding of *A* and *Y* by the covariates defined by *V*, so the weight is adjusted accordingly. In other words, the inclusion of *V* in the numerator of the weight adjusts the weight so that it is closer to 1. Weights that include the numerator portion are known as *stabilized weights*; those without the numerator portion are known as *non-stabilized weights*—i.e., numerator is equal to 1. Stabilized weights have the property that they may provide more efficient (i.e., less variable) MSM estimates(15).

An additional set of weights (given by the expression below) can be constructed for the probability of censoring which addresses potential bias of MSM estimates due to selection of the data generating process (e.g., overestimation of protective effects) due to right censoring in longitudinal analyses. Models that evaluate the probability of loss to follow-up at time *j* based on previous treatment and covariates can be fit using all subjects. Such models could be based on follow-up status (presence/absence) and observed covariates from the past. The joint product of the fitted probabilities obtained from these models could be used to represent the probability of a subject being observed in the data as of time *j*. The fitted probabilities would apply to those who were observed in the data as of *j* and applied as a weight in the regression --i.e., the subjects lost to follow up cannot contribute information at time *j* (15):

$$SWC_{ij} = \frac{\prod_{j=1}^{k} \Pr(C(j)=0 \mid C(j-1)=0)}{\prod_{j=1}^{k} \Pr(C(j)=0 \mid C(j-1)=0, \bar{A}(j-1), \bar{L}(j-1))}$$

Indeed, to account for confounding of *A* and selection bias, the weight used in the regression to estimate causal parameters, can be a joint mechanism that accounts for both of these sources of bias that is applied to an overall weight each subject at time *j*: $W_{ij} = SW_{ij*} SWC_{ij}$ (15). While IPTW estimation is one of the more straightforward estimating procedures to estimate causal effects, it is one of the most susceptible to bias, as the result of ETA violation, and, compared to other estimators, it is relatively inefficient.

Targeted Maximum Likelihood Estimation(TMLE)

As an alternative estimator to the ones discussed above, one can apply TMLE(10). The general motivation for this estimator is to target particular parameters of interest from a causal model ---*e.g., E[Y|A,W]*—which are estimated typically by maximum likelihood estimation (MLE). By doing so, one can reduce the bias/variance of the target parameters at the cost of increasing the variance of the other nuisance parameters in the model(10, 11). The TMLE has double-robust properties, similar to DR-IPTW, in the sense that if one consistently estimates either the *FX*-part or the *g*-part of the likelihood, one can obtain consistent causal estimates.

The basic idea of TMLE can be illustrated with an example from a point treatment analysis. Extension to longitudinal analyses is straightforward.

1) Obtain initial MLE estimate Q^0_n based on a regression of the model *E[Y|A,W]*. One can estimate *E[Y|A,W]* with cross-validation DSA, for example. It is not required that this initial estimate of *E[Y|A,W]* be perfectly specified.
2) Estimate $Q^1_n = Q^0_n + \varepsilon h(A,W)$ which can be implemented by a regression of Y on Q^0_n and h(A,W), where $h(A,W) = \left(\frac{I(A=1)}{g^0_n(1,W)} - \frac{I(A=0)}{g^0_n(0,W)} \right)$. This provides an estimate of ε, where ε represents a perturbation parameter whose estimate reflects residual confounding of the target parameter of the variable A by W. In general terms, ε represents the difference in the expectation of Y that is due to non-randomized assignment of treatment/no treatment left over that is not explained by the initial estimate Q^0_n.
3) Repeat Step 2, iteratively to obtain Q^*_n.
4) Calculate $Q^*_n(A=1,W)$ and $Q^*_n(A=0,W)$, and take the difference, for example, to obtain an estimate of the marginal additive risk difference for A=1 vs. A=0.

Extensions of TMLE--or Collaborative target maximum likelihood estimators (CTMLE)—have been developed which use information about Q and g "collaboratively" in the estimation procedure to: 1) reduce the number of unnecessary variables to estimate g(A,W) beyond the sufficient set necessary to control for bias, and, therefore, further reduce the variance of the target parameter estimate; and 2) avoid potential ETA violation by excluding particular confounders from g(A,W) that would most likely contribute to such a violation. Such an estimator has advantages given that excluding particular confounders to avoid ETA violation changes the analysis of a particular research question; the CTMLE overcomes this problem given the use of information Q and g parts of the likelihood. Details of these estimators are provided elsewhere for the interested reader(16).

Interpretation and Application of MSM Methodology

Overview

MSMs represent versatile tools for the examination of the examination of a variety of questions for population aging and epidemiologic research. Given these models, one can focus on particular causal parameters (marginal effects) of interest rather than conditional effects—e.g., the marginal effect of the independent contribution of a particular pathway involved in the aging process (e.g., physical activity). One can evaluate these parameters in a variety of contexts, and combine these parameters (e.g., joint contribution of parameters from separate models) to investigate processes that one may wish to examine.

For example, one could evaluate physical activity as an intervention (effect of higher than recommended levels compared with lower levels) based on observational data, and develop treatment rules that are expected to provide optimal outcomes based on the observed data (e.g., higher than recommended levels may provide greatest benefits, but may not confer benefits in particular subgroups). These rules based on observational data, in turn, could serve as potential candidate rules for examination in RCTs. In another instance, one might wish to investigate the direct mechanisms of physical activity that can affect aging outcomes—e.g., cognitive health---

independently of intermediary pathways known to affect cognition and improve with physical activity (e.g., effect of LTPA on cognitive function via pathways other than those related to hypertension and manifestations of atherosclerosis (e.g., stroke)).

In each case, the causal parameters of interest that are derived with the MSM are focused on physical activity, but they are used to address fundamentally different questions of its role in the aging process. I begin this section first with an interpretation of causal parameters based on a simple MSM for a longitudinal study of aging. I then follow this illustration with examples of how parameters from MSMs, and extensions of MSM, can be used to address more general sets of questions for aging.

Interpretation of MSMs

MSMs are amenable to the study of a variety of effects in aging, and epidemiology in general. For example, the following MSM might be formulated as one previously proposed (17) to describe the causal effect of current leisure-time physical activity (LTPA) on self-reported functional limitation (PF) for any given time *t*, in a longitudinal study where these variables are measured repeatedly over time:

$$E(Y_a(t) = 1 \mid V) = \beta_0 + \beta_1 a(t) + \vec{\beta}_2 V + \vec{\beta}_3 a(t) * V ,$$

with *V* being determined through a directed acyclic graph which encodes the hypothesized pathway model. In the model, $Y_a(t)$ might represent the treatment (LTPA)-specific PF at *t*, where PF may be a binary variable (Yes/No), and LTPA *'a'* is defined according to participation in levels of physical activity at or above the public health recommended levels (Yes/No). The causal parameter β_1 represents the change in the mean population PF, if, contrary to fact, everyone met the public health recommended guidelines for physical activity compared to no one in the population meeting those goals. The parameter $\vec{\beta}_2$ represents a vector of coefficients correspondent with the associated risks of PF with the different covariables defined by *V* that represent different subsets of the population defined at baseline (e.g., young vs. old). Lastly, the parameter $\vec{\beta}_3$, which represents a vector of coefficients also, quantifies the extent to which the different subgroups of the population defined by *V*, alter the effect of current LTPA on PF –e.g., effect of meeting public health recommended levels of LTPA in older compared to younger elderly subjects.

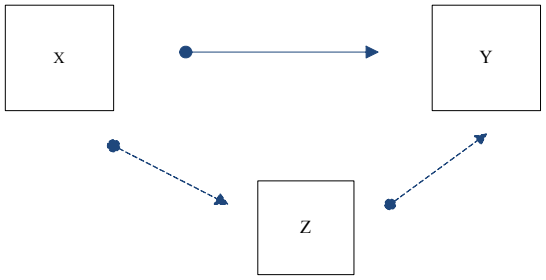
The MSM described above evaluates the current (immediate) temporal effects of LTPA (e.g., average levels of physical activity reported in previous year) on physical function status at any given time. However, other temporal effects could have as easily be described and parameterized in the given model. For example, one may not be interested in the current effects of LTPA, but interested instead in the long-term effects (or combination of short-term and long-term effects)---i.e., a 'history' of LTPA influence on functional status at any given time. One might choose, in this case, to include singular parameters that mark participation in recommended levels of LTPA at different points in time (e.g., prior to study, study baseline, current) in the model, or include a summary measure---e.g., a single parameter $\bar{a}(t)$--that represents a "smoothing" (e.g., running mean) of these parameters. The MSM causal parameter of such a measure, in the latter instance, might reflect the effect of LTPA, if contrary to fact,

everyone increased his/her average number of times they met the public health recommended levels of LTPA by 1 additional year in the preceding k years, than if he/she did not.

Other parameterized models are possible which can lend themselves to novel applications that are particularly revealing to understanding causal mechanisms in aging populations with important public health implications. In one application that is described below, MSMs were applied to examine the influence of LTPA and other factors of interest (and the temporal influences of these variables) on physical functioning patterns for an 8-year study of the elderly. Based on this work, "successful" (i.e., disablement-free or postponed disablement) and "unsuccessful" patterns of aging were identified and evaluated probabilistically with respect to hypothetical patterns of LTPA (counterfactuals) that could have occurred study population.

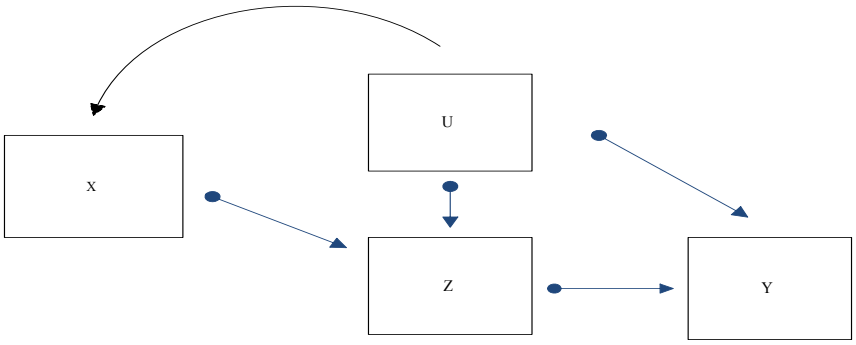
Application of MSMs: Estimation of direct causal effects

Situations may arise when one may not only be interested in a 'total' causal effect, but also in its partition into direct effects and indirect effects on a given outcome. For example, an effect of physical activity with respect to a particular outcome (cognitive health) might be partitioned into its potential direct (e.g., increased brain-derived neurotrophic factor, preservation of frontal-cortical volume) and indirect effects (e.g., improved cardiovascular health), each of which could be estimated separately. This can be illustrated in the following graph.



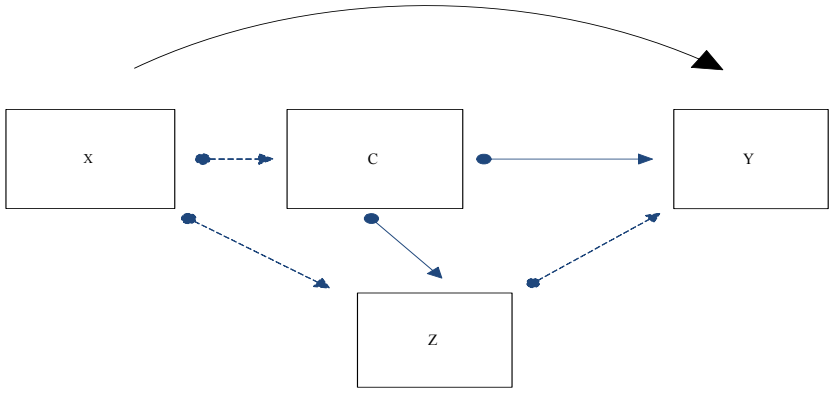
Here, one may wish to quantify the direct effect of X on Y (solid line) independent of the pathway through the causal intermediate Z (dashed line). While quantification of this direct effect may be of interest, it can be problematic to implement with standard analytical approaches.

Pearl notes that analytical methods that examine the joint distributions of variables to isolate direct effects are susceptible to 'back-door' effects as illustrated with the next graph(1).



By conditioning on the causal intermediate Z, and given the association of X and Z, a spurious association between X and Y could form by way of a 'back-door' effect (represented by curvilinear arrow) through unmeasured confounder U. U is a potential confounder that is not measured. Pearl advises that the best approach in this situation is to 'physically' remove the influence of X on Z (which is possible with structural equation modeling or MSMs), which would then allow one to test for a direct effect between X and Y(1).

Petersen, *et al*. show how MSMs are particularly suitable for quantifying effects in the presence of causal intermediates. In the following graph, estimation of the direct effect of X on Y is complicated by the occurrence of other variables on the causal pathway:



In the graph, X affects C, which, in turn, affects Z and Y. Estimation of a direct effect of X on Y (curvilinear line) in this instance is equivalent to the estimation of effects in the presence of time-dependent confounders. By excluding C, an essential confounder of Z and Y is missing, which can result in a biased estimate of the direct effect of interest. Inclusion of C, however, could 'adjust away' the direct effect. Therefore, alternative methods (e.g., MSMs) are needed to obtain unbiased estimates of interest in these situations(18).

Given that appropriate methods are applied and confounding is properly addressed, direct effects can be obtained by the evaluation of different counterfactuals. For example, one could estimate the direct effect of x` vs. x by:

$$E[Y_{x'z}] - E[Y_{0z}]$$

which estimates the direct effect of x on the counterfactual distribution of Y (e.g., walking speed) by taking the mean difference in Y if, contrary to fact, everyone in the population experienced x` (e.g., high LTPA) compared with if everyone experienced the absence of x`, with the intermediate variable fixed at some level Z=z (e.g., leg strength). This estimate could be obtained directly from a MSM model that, based on the above example, would include causal parameters for X and Z.

Alternatively, one want to examine the direct effect of x` (again, for example, high LTPA) but in a population with observed distribution of LTPA, defined as x, to assess the additional direct effects on Y if, contrary to fact, everyone in the population participated in high LTPA:

$$E[Y_{x'z}] - E[Y_{xz}]$$

This could be derived, with some manipulation, based on an estimated MSM as mentioned before. Such effects are based on population-intervention MSMs, which can be used to investigate distributions of counterfactual outcomes based on some population-level intervention (e.g., increase of 30 minutes of physical activity-5 days/week) relative to observed (crude) outcome distributions present already in the population(19).

The direct effects shown above are examples of *controlled direct effects (CDE)*. That is, the intermediate variable Z is fixed at a given level, and the direct effect of X is evaluated under that condition. In other applications, one may want to assess the direct effect of X in the presence of Z, when Z remains at its observed levels in the population. In this case, the direct effect is known as a *natural direct effect (NDE)*. NDE can be derived from CDE by weighting the CDE by the distribution of Z observed in the population conditional on the observed X(1):

$$\sum_z E[Y_{x'z}] - E[Y_{xz}]P(Z = z \mid X)$$

Further details and examples of these different types of effects are provided elsewhere(1, 18).

Extension of MSMs: History-adjusted MSMs (HAMSMs)

HAMSMs represent a generalization and extension of MSMs for the evaluation of causal effects with longitudinal data(20, 21). Given HAMSMs, one can specify a model by which causal effects of interest are updated over time with respect to time-varying covariates—that is, causal effects can be examined <u>*explicitly*</u> with respect to their potential alteration by changes in a population's covariate distribution over time. By contrast, traditional MSMs allow one to examine causal effects, which may also be updated over time given changes in covariate patterns in the population, but for which one cannot model the influence of these time-varying covariates on the effects in question(20) . In other words, with traditional MSMs, one obtains a causal effect which is a composite, summary measure of effects that may differ by time-varying covariates (e.g., incident cardiovascular disease) in the population.

HAMSMs operate similarly to traditional MSMs which have been discussed until now. HAMSMs model the conditional distribution of treatment-specific counterfactual outcomes, but with the following difference: in HAMSMs, at each timepoint *j*, baseline status is updated and the effect of future treatment (relative to time *j*) is evaluated conditional on covariates updated through time *j* (20). For example, the following HAMSM could be assumed as a mean structure

for the expected effects of LTPA measured over the interval *(j, j+m)* on mini-mental state exam scores (MMSE) measured at the end of these intervals, given differences in subpopulations defined by covariates $V(j)$ up to time *j*:

$$E(Y(j+k)_{\bar{A}(j-1),\underline{a}(j)} \mid V(j)) = \beta_0 + \beta_1 \underline{a}(j) + \beta_2 V(j) + \beta_3 \underline{a}(j) * V(j), \quad j=0, 1,\ldots,k$$

The counterfactual in this instance $Y(j+k)_{\bar{A}(j-1)\underline{a}(j)}$ denotes the treatment-specific MMSE that is indexed by observed LTPA history up to but not including time *j*: $\bar{A}(j-1) = (A(0),\ldots,A(j-1))$, and a future counterfactual LTPA regimen $\underline{a}(j)$ which is defined here over the interval *(j, j+m)* where *m=1*, for example. Such a model might be appropriate for the examination of short-term effects of LTPA on decline in MMSE—a marker of cognitive decline that may be sensitive to short-term physiologic changes that are associated with normal aging (e.g., glucose intolerance, hypertension). The estimated coefficients from this model could be informative with respect to potential differential effects of short-term changes in LTPA on decline of MMSE in light of subgroups with different histories of hypertension and glucose intolerance over the course of a given study. Although a parametric HAMSM is illustrated in the example above, nonparametric approaches have been applied in situations where parametric models were unfeasible(22).

Estimation of HAMSMs corresponds to the simultaneous fitting of *j*- time point specific MSMs, in which treatment and covariate history up to each time point are treated as baseline covariates, and the dependence of the counterfactual outcome on the future treatment is modeled(20). Estimators of MSMs have been adapted for HAMSMs (e.g., G-computation, IPTW, Double-Robust IPTW, TMLE). Details of these different estimators are given elsewhere(10, 20, 21).

Lastly, to illustrate the utility and potential advantages that these models have for epidemiologic research related to aging, as well clinical medicine (e.g., HIV treatment), I describe, based on recently developed methods, how the estimates that are derived from HAMSMs can be used to examine *dynamic treatment regimens* based on observed data (23). Dynamic treatment regimens represent the individualized set of treatment rules used by a physician, given his/her patient's evolving medical status, to aid in the success of the patient's outcome. Based on HAMSMs, at each time point *j*, one can potentially estimate the optimal treatment effect given different subsets of the covariates one conditions on as of time *j* (23). This is analogous to carrying out a RCT, where subjects may be initially stratified according to their covariates (e.g., age, function). In this instance, the classic RCT would identify the optimal treatment at the given time point. Hypothetically, however, the trial could be repeated for each time point in the study (subjects repeatedly stratified according to time-dependent covariates and randomized to treatment) with the end result that this series of trials would identify, for each time point, the future treatment which could provide the best expected outcome given subjects' covariate history. Petersen, *et al*. examine the application of HAMSMs to develop individualized treatment rules for patient populations(23).

In summary, most studies to date have not been able to address adequately fundamental questions related to population aging because of limitations of the analytical techniques and/or study designs that have been available to researchers. MSMs and HAMSMs, and causal statistical methods in general, represent an advancement over these traditional methods that allow greater flexibility with respect to research questions that can be addressed and the

examination of causal effects; the utilization of data available for analysis; and improved control of confounding factors and assessment of response heterogeneity.

References

1. Pearl J. Causal inference in statistics: An overview. Statistics Surveys 2009;3:96-146.
2. Robins JM. Association, causation, and marginal structural models. Synthese 1999;121:151-79.
3. van der Laan MJ, Robins J. Unified Methods for Censored Longitudinal Data and Causality, 2002.
4. Greenland S, Robins J, Pearl J. Confounding and Collapsability in Causal Inference. Statistical Science 1999;14:29-46.
5. Wang Y, Petersen M, Bangsberg D, van der Laan M. Diagnosing Bias in the Inverse Probability of Treatment Weighted Estimator Resulting from Violation of Experimental Treatment Assignment. Berkeley Electronic Press/UCB Biostatistics Division 2006;211.
6. Bembom O, van der Laan MJ. A practical illustration of the importance of realistic individualized treatment rules in causal inference. Electron J Stat 2007;1:574-596.
7. Neugebauer R, van der Laan M. Locally Efficient Estimation of Nonparametric Causal Effects on Mean Outcomes in Longitudinal Studies. Berkeley Electronic Press/UCB Biostatistics Division 2003;134.
8. Robins J. Robust estimation in sequentially ignorable missing data and causal inference models. American Statistical Association. Alexandria,Va, 1999:6-10.
9. Neugebauer R, van der Laan M. Why prefer double robust estimators in causal inference? Journal of Statistical Planning and Inference 2005;129:405-426.
10. van der Laan M, Rubin D. Targeted maximum likelihood learning. International Journal of Biostatistics 2006;2.
11. Gruber S, van der Laan M. Targeted Maximum Likelihood Estimation: A Gentle Introduction. Berkeley Electronic Press 2009;252.
12. van der Laan M, Dudoit S. Unified Cross-Validation Methodology For Selection Among Estimators and a General Cross-Validated Adaptive Epsilon-Net Estimator: Finite Sample Oracle Inequalities and Examples. Berkeley Electronic Press/UCB Biostatistics Division 2003;130.
13. Sinisi S, van der Laan M. Loss-Based Cross-Validated Deletion/Substitution/Addition Algorithms in Estimation. Berkeley Electronic Press/UCB Biostatistics Division 2004;143.
14. Wang Y, Bembom O, van der Laan M. Data Adaptive Estimation of the Treatment Specific Mean. Berkeley Electronic Press/UCB Biostatistics Division 2004;159.
15. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550-60.
16. van der Laan M, Gruber S. Collaborative Double Robust Targeted Penalized Maximum Likelihood Estimation. Berkeley Electronic Press 2009;246.

17. Tager IB, Haight TJ, Yu Z, Sternfeld B, van der Laan MJ. Longitudinal study of the effects of physical activity and body composition on functional limitation in the elderly. Epidemiology 2004;15:479-493.
18. Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. Epidemiology 2006;17:276-84.
19. Hubbard AE, van der Laan MJ. Population intervention models in causal inference. Biometika 2008;95:35-47.
20. van der Laan M, Petersen M, Joffe M. History-Adjusted Marginal Structural Models and Statically-Optimal Dynamic Treatment Regimens. International Journal of Biostatistics 2005;1.
21. Petersen M, Deeks S, Martin J, van der Laan MJ. History-adjusted Marginal Structural Models for Estimating Time-varying Effect Modification. American Journal of Epidemiology 2006;166:985-993.
22. Bembom O, van der Laan MJ, Haight T, Tager I. Leisure-time Physical Activity and All-cause Mortality in an Elderly Cohort. Epidemiology 2009;20:424-430.
23. Petersen M, Deeks S, van der Laan M. Individualized Treatment Rules: Generating Candidate Clinical Trials. Statistics in Medicine 2007;26:4578-4601.

Chapter 3

Effects of Body Composition and Leisure-Time Physical Activity on Transitions in Physical Function

Introduction

Long-term effects of different aging factors on physical disablement deserve consideration in aging research. Researchers may wish to know how different patterns of these factors affect functional patterns over time. Such functional patterns are of interest: functional patterns indicative of freedom from functional limitation until the end of life are clearly more advantageous than functional patterns characterized by physical limitation and dependence.

Functional status in the elderly does not follow a steady course of decline, but represents more episodic transitions into and out of physical limitation (1). Therefore, intrinsic to any description of how aging factors might influence functional patterns is the role of previous functional status on future functional status. The question becomes how to evaluate factors that may influence the course of disablement given the effects of previous function on these factors and on future function. Another question is: how can one expand a study to examine these factors over time and examine their long-terms effects on differences in patterns of function and disablement.

Body composition and leisure-time physical activity have been observed to play an important role in the disablement process in the elderly (2-6). The disablement process is a model that depicts a pathway of events in aging adults that marks the gradual decline in bodily systems and links healthy functioning with functional limitation, disability, and mortality. As formulated, the process begins with underlying patho-physiological abnormalities that develop into structural abnormalities in specific body systems (impairments). These impairments can impact a person's ability to function normally, with the potential consequence of a person's inability to interact with and meet the demands of their social/physical environment (disability) (2). Recently, Stewart suggested refinement of the disablement model to include physiological changes that accompany normal aging that contribute, independently of pathological pathways, to the decline in bodily systems and physical fitness levels that lead to functional limitation and disability (7).

Previous studies have elaborated on the underlying risk factors (biological, demographic, social, environmental attributes) that influence the underlying physiology and subsequent events along the pathway to disablement (2) (8) (7) and the complex network of intermediary factors that contribute to the disablement process (3). Of particular interest is the role of physical activity in the prevention of diseases that could lead to limitation, or to reduction in the occurrence of functional limitation in healthy elders (9-13). Additionally, studies based on surrogates of underlying physiological processes in the disablement process (measures of obesity in relation to impaired insulin metabolism) attempt to shed light on the roles and the potential mitigating effects of physical activity on the progress of disablement. (5) (14).

Recently, effects of physical activity (LTPA) and the ratio of lean body mass to fat mass (L/F) were examined for associated and causal effects on self-reported functioning. In a cross-sectional analysis (4), L/F was associated with faster walking speed and improved functioning.

Moreover, a relative measure of lean to fat mass, rather than lean body mass alone, was the more important factor related to physical performance and physical functioning. A subsequent longitudinal analysis evaluated the combined causal effects of L/F and LTPA on physical functioning (5); L/F had a greater effect than LTPA on reduction in the log-odds of functional limitation, except in the presence of obesity. Those studies suggested that the beneficial effects of physical activity were most likely mediated through reduction of fat mass relative to lean mass. In addition, in the presence of obesity, it was found that improvement in muscle mass had little effect on the preservation of functioning (5).

Descriptions of transitions between various levels of physical function and disability in elderly subjects have been reported (1, 15-19)**.** Beckett and colleagues showed that the decline in mean level of physical function in their elderly subjects did not imply that all subjects followed a steady course of decline; some subjects recovered from disability even at the oldest ages (15). Anderson *et. al*. emphasized the heterogeneity of transitions between states of disability that can occur in elderly subjects, and demonstrated the importance of the incorporation of knowledge about previous patterns of disability for estimation of the probability of subsequent transitions in disability status (1). Other studies (16-19) emphasize the prognostic importance of prior disability episodes on new episodes, and the high rates of recovery from disability that older subjects experience. The implication of the later findings supports the argument that additional efforts could extend the recovery and independence of older subjects who are at high risk of recurring disability(17, 19).

The present investigation was undertaken to extend current understanding of transitions between states of functioning in the elderly within the context of the disablement model through the application of Marginal Structural Models (MSMs) (20). Estimation of these models, given the causal analytical framework discussed in Chapter 2, provides less biased estimates of the effects of interest in the presence of the time-dependent confounding that is inherent in the disablement model and permit more direct population-level inferences than can be derived from more conventional statistical approaches (21) (5). The application of statistical methods for causal inference provides the opportunity to unravel the complex *causal* relationships in the disablement model to an extent not easily achieved or even possible with standard statistical methods.

Materials and Methods

Study Sample

Eligible subjects were 947 women and 708 men (n=1655) who were a subset of 2092 men and women ∃ 55 years who resided in and around Sonoma, CA and who were participants in a community-based, longitudinal study of the effect of aging on functioning in the elderly (22) (23). The protocols were approved by the Committee for the Protection of Human Subjects, University of California, Berkeley and the Committee for Human Research at the University of California, San Francisco.

The 1655 subjects were those who had bioelectric impedance measurement, physical performance and physical function data at the baseline evaluation (May, 1993 – December, 1994). Full details of the protocol have been published previously (22-24). Data from the present analysis were derived from the baseline and three subsequent evaluations (September, 1995 – November, 1996; June, 1998 – October, 1999; February, 2000-March, 2001). Complete data

from all four evaluations and from the first three and two evaluations were available for 648**,** 884, 1236 subjects, respectively.

Assessment of functional limitation and physical performance

Self-reported functional limitation was based on 10 questions (See Appendix 2) that assessed the degree of difficulty that a participant reported in various domains of physical functions (upper- and lower-body domains of varying complexity) (4) (25) (26, 27 , 28).

Participants who reported "a lot of difficulty" doing one or more of the functions or not doing at least one function because they were unable or were advised by a doctor not to do so were classified as having self-reported limitation.

Measurement of body composition

Estimates of fat mass and lean mass were derived from resistance and reactance measured by bioelectric impedance (BIA) (BIA101Q Quantum Body Composition Analyzer System , RJL Systems, Clinton Township, MI) based on study-specific validation equations (4). The total variance in lean mass accounted for by these regressions was 0.85 and 0.80 for males and females, respectively. Similar regressions were performed for lean plus bone mass. Fat mass (kg) was obtained by subtraction of lean plus bone mass from body weight. Lean-to-fat ratio (L/F) was defined as the ratio of lean mass to fat mass (both in kg). Measurements of appendicular fat-free mass were not available.

Measurement of leisure time physical activity (LTPA)

At each evaluation, subjects reported their average weekly participation over the past 12 months (average number of times per week for each activity) in 22 specific physical activities that spanned a wide range of energy expenditures (23). Each activity was assigned a MET value (1 MET ≅ oxygen consumption of 3.5 ml/kg min^{-1}) from a standard compendium of MET values (29). Two separate classifications of LTPA were created.

A continuous LTPA variable was derived based on a weighted sum of the frequency and MET values of the 22 activities for each subject $\sum_i \left(\#times/week * Mets\right)_i \; i = 1,...,22$ (23). Based on recommended levels of physical activity (30), we created a four-level categorical variable that corresponded to no LTPA, (0 METs/week), insufficiently active (0-22.5 METs/week, where 22.5 METs/week represents the minimum recommended level based on the MET value for brisk walking (4.5)), meeting minimum recommendation (22.5-35 METs/week), and highly active ($\geq$ 35 METs/week).

At baseline, subjects were asked if the level of their physical activity had changed in the preceding five to ten years. Responses were recorded into a dichotomous variable (1=report of any decline; 0=no change or increase).

Other covariates

Weight and height were measured by standard protocol at each visit, and body mass index (BMI) was calculated as weight (kg)/height $(m)^2$. We categorized BMI into three groups (<25, $25-<30$, $\exists 30$ kg/m^2) (31, 32). At each survey, subjects were classified as having none, 1 or >1 chronic diseases based on the new or past occurrence of self-reported cancer, cardiovascular

disease, cerebrovascular disease, diabetes mellitus, kidney, liver, or Parkinson's disease. The presence of depression was based on a score of 16 or greater on the CES-D depression scale (33) and current use of an anti-depressant medication (direct inspection of all medications at each interview). Smoking at each survey was classified as never, current, and former (23). Subjects rated their overall health (excellent, good, fair, poor; summarized as a dichotomous variable: 0=excellent/good, 1=fair/poor). Living arrangements was defined for each subject as living alone, living with a spouse, living with others (34). Walking speed (ft/s) was measured by the number of feet walked in 60 seconds (23).

Statistical Analysis

Description

Functioning in the elderly was evaluated as a stochastic process (35) which changes over time. For example, $Y(t) \in (0,1), t = 1,...4$ can be viewed as a discrete-time process during which a person might or might not exhibit signs {Y(t)=1, Y(t)=0, 1=limitation 0=no limitation} associated with functional limitation at observation times *t*. We can represent this stochastic process with a multivariate vector $\overline{Y} = \overline{y}$ [Y(1)=y(1),Y(2)=y(2),Y(3)=y(3),Y(4)=y(4)]. Thus, we can evaluate the process as different "histories" or "courses" of disablement (transitions) defined on an interval (*t=1,…,4*) and having a distribution, $\Pr(\overline{Y} = \overline{y})$, and state-space Ω=16 possible realizations (2^4 combinations of impaired/not impaired over 4 time points). We are interested in how the distribution $\Pr(\overline{Y} = \overline{y})$ might vary, given underlying factors that occur in the disablement process that can lead to different transitions, in particular, the marginal (unconditional) effects of L/F and LTPA.

We applied MSMs, and estimated these models, as in previous studies (36-38), to evaluate the effects of exposure variables (e.g. L/F, LTPA) in the presence of time-dependent confounders (e.g. walking speed, health status) which are affected by previous levels of the exposure variables. Standard approaches, under these circumstances, would yield biased estimates of the effects (39). The covariates described above are assumed to represent all the measurable confounders of LTPA and L/F. In the MSM analytical framework, we control for the effects of these confounders relative to L/F and LTPA through time-specific exposure ("treatment") models. In addition, we can control for selection bias from various sources of loss-to-follow-up by use of censoring models at each time point. From these respective models, we obtain subject-specific, time-specific weights (20), which can be applied as estimators of the MSM to obtain unbiased estimates of the marginal effects of L/F and LTPA.

Application of Marginal Structural Model (MSM)

MSMs are based on the concept of counterfactual variables. An exposure-specific counterfactual (potential outcome) variable is a random variable that represents a subject's set of outcomes had the subject, contrary to fact, had an exposure history other than the one actually observed (20). Formal presentation of the theory, assumptions and applications of MSMs toward epidemiological research have been reported extensively (20) (36-39) and summarized in a preceding chapter (i.e., Chapter 2). A formal data analysis which assessed the marginal effects of L/F and LTPA on functional limitation is available (5).

We assume a generalized MSM for the marginal distribution of $\overline{Y}_{\bar{a}}$:

$$\Pr(\overline{Y}_{\bar{a}} = \bar{y}) = m(\bar{a}, \bar{y} \mid \beta)$$

$\overline{Y}_{\bar{a}}$ is defined as the multivariate vector that represents different transitions , or histories of functioning if, contrary to fact, subjects followed joint exposure history $\bar{a} = (\overline{LTPA}, \overline{L/F}) = (LTPA(0), L/F(0), ..., LTPA(t), L/F(t))$. We assume $\overline{Y}_{\bar{a}}$ has a joint probability distribution:

$$\Pr(\overline{Y_{\bar{a}}} = \bar{y}) = \Pr(Y_{\bar{a}}(1) = y(1), Y_{\bar{a}}(2) = y(2), Y_{\bar{a}}(3) = y(3), Y_{\bar{a}}(4) = y(4)).$$

To formulate the MSM model, we factorize $\Pr(\overline{Y_{\bar{a}}} = \bar{y})$ as the product

$$\prod_{t=1}^{4} \Pr(Y_{\bar{a}}(t) = y(t) \mid \overline{Y_{\bar{a}}}(t-1) = \bar{y}(t-1))$$

We model each term $\Pr(Y_{\bar{a}}(t) \mid Y_{\bar{a}}(t-1))$ of the product separately at each *t*, where we assume a logistic MSM that evaluates the exposure-specific mean functional limitation as a linear combination of L/F and LTPA history, $\bar{a}$, through *t* within strata of past functional history $\bar{y}(t-1)$:

$$Logit \quad \Pr(Y_{\bar{a}}(t) = 1 \mid \overline{Y}_{\bar{a}}(t-1) = \bar{y}(t-1)) = \beta_{0t} + \beta_{1t} * \bar{a}(t) + \beta_{2t} * \bar{y}(t-1)$$

(See Appendix 2 for the exact formulation of $\bar{a}(t)$ and $\bar{y}(t-1)$ employed in the models. Note when *t=1*, $\bar{y}(t-1)$ is an empty set.) The evaluation of the effects of $\overline{LTPA}$ and $\overline{L/F}$ on $Y_{\bar{a}}$ is based on the assumption that, for any given *t*, LTPA affects L/F, and both of these affect Y: LTPA is based on average activity levels reported by a subject over the 12 months prior to *t*; L/F is measured approximately at *t*; and Y is a measure of the functional limitation reported by a subject for the month prior to *t*. Given the observed changes in L/F over the study period were small, we assume any differences in L/F between the period in which L/F affects Y(t) and L/F is measured at *t* are negligible. The parameters (β_{1t}, β_{2t}) have no direct interpretation and are not reported; rather, we utilize these parameters, as described in the next section, to construct the distribution of interest $\Pr(\overline{Y}_{\bar{a}} = \bar{y})$. The parameters are estimated by solution of the Inverse Probability of Treatment Weight (IPTW) and Inverse Probability of Censoring Weight (IPCW) estimating equations (20). Solution of these estimating equations is the equivalent of a weighted regression with $Y(t)$ as a function of $\bar{a}$ *(t)* and past functioning history $\bar{y}(t-1)$, with logit link and subject-specific, time-specific weights. A general discussion of the calculation of subject-specific weights can be found in the preceding chapter on general methods and in various articles (20) (39). A practical example of the application of subject-specific weights in a previous analysis using MSMs is available also (5).

Regressions were implemented with standard logistic regression software (Proc Logistic, SAS Version 8.2) with weights.

Computation of Counterfactual Transition Distributions

To compute the distribution of transitions in functioning $\Pr(\overline{Y_{\bar{a}}} = \bar{y})$ for a counterfactual exposure history, $\bar{a} = (\overline{LTPA},\ \overline{L/F})$, we first estimate the marginal distributions of functional status $\Pr(Y_{\bar{a}}(t) = y(t) \mid \overline{Y_{\bar{a}}}(t-1) = \bar{y}(t-1))$ for each time *t* based on the time-specific parameter estimates β_{1t}, β_{2t}; second, we compute the product of these marginal distributions of functional status from *t=1,..,4*:

$$\prod_{t=1}^{4} \Pr(Y_{\bar{a}}(t) = y(t) \mid \overline{Y_{\bar{a}}}(t-1) = \bar{y}(t-1)$$

Details of this computation are given in Chapter Appendix 2.
Confidence intervals (95%) for each transition probability were calculated with the standard error based on a bootstrap distribution of 1000 estimates for each of the different transitions. If the distributions were not normal, the 2.5 and 97.5 percentiles of the distribution were selected.(40)

Results

Baseline characteristics of the study population are displayed in Table 1. The median (IQR) duration of follow-up for both females and males was approximately 6.4 years (5.6-6.9). The median number of activities for which subjects reported difficulty was 2.0 (IQR 1-4 in women, 1-3 in men), with the three most frequently reported activities being: 1) getting up from a stooped, kneeling or crouched position; 2) stooping crouching, or kneeling; and 3) lifting items greater than 10lbs (women) and standing for longer than 15 minutes (men).

Tables 2 and 3 present proportions of transitions that represent the onset of functional limitation without recovery in women and men, respectively, along with five of the many possible counterfactual, joint exposure histories for L/F and LTPA. The proportions of transitions based on the observed data (column 3) represent the counterfactual distribution actually observed in the study population. Fifty percent of women and 71 percent of men did not report any limitation over the entire study period. Approximately nine percent of women and 2.4 percent of men reported limitation in at least one function at all four surveys. Onset of functional decline without recovery, (summation of the proportions of the other transitions in tables 2 and 3), was observed for 16.3 percent of women and 9.3 percent of men. Table 4 presents proportions of transitions which represent full recovery from a baseline limitation and other transitions which represent temporal (non-monotonic) patterns of limitation and recovery over time. Approximately five percent of women and 3.4 percent of men experienced full recovery from baseline limitation (rows 1-2, column 3), whereas 14 percent of women and 11.7 percent of men experienced a new limitation followed by recovery (rows 3-4, column 3).

The proportions of all possible transitions are re-distributed if the study population, contrary to fact, experienced the selected counterfactual exposure histories (Tables 2-4, columns 4-8). For example, we generally observe an increase in risk of functional impairment in the

subjects if they did not exercise over the study period but maintained the same levels of L/F (Tables 2-4, column 4). The proportion of women without any limitation during the study (Table 2, row 1) drops from 50 percent to 38.4 percent, while the proportion who are limited at all time points rises from approximately nine percent to 15.3 percent. A similar pattern occurs for the men. Proportions of transitions suggestive of functional limitation were not greater in all cases and were not statistically different from the proportions that were actually observed in the data (column 3).

If contrary to fact, subjects exercised at high aerobic levels (>=35 Mets/week) at all surveys, the likely onset of functional limitation without recovery would be expected to decrease (Tables 2, 3 column 5). We observed a reduction of 53 percent (7.4 → 3.5 percent) in women's chances of development of a functional limitation at the last survey compared with their actual observed history (Table 2, row 2, column 5 vs. column3). Furthermore, high, sustained levels of aerobic exercise play a role in the delay of onset of functional limitation (Table 2, rows 3 and 4). For example, the risk of women who experience functional decline at the 2nd and 3rd surveys is reduced approximately 38 percent and 45 percent, respectively (4.7 → 2.9 percent 2nd survey, 4.2→2.3 percent 3rd survey). Although a reduction of late-stage limitation occurs in the men with increased levels of exercise, functional decline does not appear to be reduced at earlier stages of the process (Table 3, rows 3 and 4). High levels of physical activity improved the chances of full recovery in subjects with limitation at baseline (Table 4, rows 1 and 2, column 5 vs. column 3). For women, the percent increase in recovery for higher levels of physical activity over levels of physical activity in the observed data was 43.8 percent (6.9 vs. 4.8 percent); in men the increase was one-fold (7.1 vs 3.4 percent). The estimates used to calculate these percent increases were imprecisely estimated (column 5); therefore, the increases are not statistically significant. Nonetheless, there is an indication that recovery from functional limitation would increase with higher levels of physical activity.

Lower risk of functional decline and higher chance of recovery from functional limitation are not solely attributable to consistent participation in high levels of LTPA. If women were initially sedentary at baseline, but increased their levels of physical activity over time (table 2, column 7), with the likely associated benefits of higher L/F, their chances of functional decline would not differ significantly than if they maintained high levels of LTPA (table 2, column 5) throughout the study. In men, a similar comparison of these two exposure histories suggest a further reduced risk of functional limitation when L/F levels increase with increasing exercise over time (table 3, column 7 vs. column 5). In the counterfactual world where the population experienced a faster rate of decline in their physical activity (column 8) and, consequently, a lower L/F, the risk of functional decline would be as large as if they did not exercise at all.

We also examined the distribution of functioning over time, if subjects were not physically active, but if contrary to fact, their L/F was one unit greater than their observed L/F (column 6). The effects appear to vary somewhat for men and women. In women (table 2), we observed an increase in the proportion of subjects without functional limitation (row1, column 3 vs. column 6, 50.5 →. 63.9 percent) and subjects who experience a delayed functional limitation (row 2, 7.4→10.1 percent), although these are both imprecisely estimated. In addition, we observed significant reductions in the proportions of women subjects who experienced limitation a majority of time during the study period (rows 3-5: 4.2 →2.1 percent; 4.7→1.2 percent; 8.7→ 3.0 percent). The benefits of higher L/F in the absence of physical activity with respect to function occur for men but to a lesser extent than in women (table 3). The proportion of men living without functional limitation during the study increased to 73 percent (table 3, row 1, column 6) as a result of there being fewer men who experience onset of functional limitation (table 3, rows 2-4, column 6).

Table 1.: Characteristics of Study Population at Baseline Interview				
	Body Composition Data Available (n=1655) *****		**Body Composition Data Not Available** (n=437)	
Variable	**Females** (947)	**Males** (708)	**Females** (299)	**Males** (138)
Age†	69.0 (56-84)	69.5 (56-82)	75.0 (58-89)	75.0 (56-89)
Body Mass Index (kg/m^2)	25.5 (20.3-34.8)	26.8 (22.0-34.4)	25.8 (19.2-35.6)	26.6 (21.2-37.7)
Lean-to-Fat Ratio	1.45 (0.98-2.45)	2.6 (1.9-4.2)	Not measured	Not measured
Walking Speed (feet/second)	2.3 (1.3-3.0)	2.3 (1.5-3.0)	2.0 (1.0-2.8)	1.8 (1.2-2.8)
METS/week of Physical Activity				
0	6.6%	3.7%	20.7%	15.2%
0-<22.5	22.8%	18.8%	20.7%	18.1%
22.5-<35	20.0%	14.4%	15.7%	21.7%
=>35	50.4%	63.1%	38.5%	43.5%
missing	0.2%	0.1%	4.3%	1.4%
Chronic Health Condition				
No	57.6%	53.2%	44.6%	38.4%
Yes	42.4%	46.8%	55.4%	61.6%
Depression‡				
	81.0%	90.2%	11.0%	6.5%
Missing	15.0%	6.4%	73.6%	76.8%
	4.0%	3.4%	15.4%	16.7%
Decline in Physical Activity Prior to Baseline				
None	62.8%	66.6%	44.5%	44.9%
Decline	36.7%	33.2%	54.8%	55.1%
Missing	0.4%	0.3%	0.7%	--

Living Arrangements				
Lives alone	39.7%	11.6%	46.0%	15.2%
Lives with spouse	53.1%	83.6%	40.6%	79.0%
Lives with non-spouse	7.2%	4.8%	13.4%	5.8%
Smoking History§				
Never	48.9%	33.6%	49.5%	29.0%
Current	8.7%	6.5%	10.0%	10.9%
Ex	42.4%	59.9%	40.5%	60.1%
Self-Rated Health				
Excellent/Good	86.1%	84.6%	68.4%	64.5%
Fair/Poor	13.6%	15.4%	30.6%	34.8%
Missing	0.3%	0.1%	0.9%	0.7%
Self-Reported Functional Limitation				
None	70.9%	84.7%	50.7%	58.7%
Any	29.1%	15.3%	49.3%	41.3%
ADL Abnormality§				
None	97.7%	98.7%	82.2%	80.4%
Any	2.3%	1.3%	17.8%	17.4%
Missing	--	--	--	2.2%

***** 104 subjects could not be used from the baseline data for analysis due to missing data on 1 or more covariates

† All continuous variables are expressed as median (5^{th}-95^{th} percentiles)

‡ Based on CES-D score and use of anti-depressant medication

§ Not included in treatment or censoring models due to lack of association with outcome (smoking) or numbers too small (ADL abnormality).

Table 2. Proportions of Selected Transitions for Onset of Functional Limitation Over Surveys 1-4 (s1-> s4) in Women.

Transitions of Functional Limitation (Yes=1,No=0)	**N**	**Observed*** **Data S1,…S4**	**Selected Counterfactual Exposure Histories of Lean:Fat Ratio (L/F) and Physical Activity (LTPA)†**				
			Median L/F, No LTPA all surveys	**Median L/F, High LTPA all surveys**	**1 Unit over Median L/F, No LTPA all surveys**	**Increasing LTPA, L/F‡ s1->s4**	**Decreasing LTPA, L/F§ s1->s4**
(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
0 0 0 0	**192**	**0.505**	**0.384**	**0.490**	**0.639**	**0.481**	**0.310**
			0.244 , 0.524†	**0.427, 0.553**	**0.395, 0.783**	**0.368, 0.594**	**0.177, 0.443**
0 0 0 1	**28**	**0.074**	**0.127**	**0.035**	**0.101**	**0.039**	**0.129**
			0.031, 0.223	**0.006, 0.064**	**0.022, 0.366**	**0.008, 0.107**	**0.047, 0.257**
0 0 1 1	**16**	**0.042**	**0.063**	**0.023**	**0.021**	**0.016**	**0.107**
			0.011, 0.115	**0.005, 0.041**	**0.005,0.069**	**0.003, 0.050**	**0.037, 0.177**
0 1 1 1	**18**	**0.047**	**0.038**	**0.029**	**0.012**	**0.026**	**0.053**
			0.005, 0.071	**0.009, 0.049**	**0.003, 0.041**	**0.009, 0.062**	**0.019, 0.087**
1 1 1 1	**33**	**0.087**	**0.153**	**0.131**	**0.030**	**0.134**	**0.172**
			0.069, 0.237	**0.089, 0.173**	**0.011, 0.068**	**0.058, 0.210**	**0.088, 0.256**

* Proportions of transitions based on distribution (i.e. frequencies) of occurrence in the observed data. The "observed" category, in fact, is one of the set of all possible counterfactual exposure histories. Proportions of transitions, under other counterfactual exposure histories, computed based on estimates from MSM models.

† Confidence intervals that overlap across different exposure histories for given transitions indicate no statistical difference at alpha 0.05 level. Confidence intervals derived from the standard error of a distribution of 1000 bootstrap estimates for each transition probability.

‡ Starting with lowest LTPA category at baseline, there is a 1 unit increase in LTPA category with each successive survey. In conjunction with this increase in LTPA is a lag increase L/F over the population median level (+0.2, +0.5) at time=3, 4.

§ Beginning with the highest LTPA category at baseline, there is a 1 unit decrease in LTPA category with each successive survey. There is a lag decrease in L/F below the population median (-0.2,-0.5) at time=3, 4.

Table 3. Proportions of Selected Transitions for Onset of Functional Limitation Over Surveys 1-4 s1-> s4 in Men.

Transitions of Functional Limitation Yes=1,No=0	N	Observed* Data S1,…S4	Selected Counterfactual Exposure Histories of Lean:Fat Ratio L/F and Physical Activity LTPA†				
			Median L/F, No LTPA all surveys	Median L/F, High LTPA all surveys	1 Unit over Median L/F, No LTPA all surveys	Increasing LTPA, L/F‡ s1->s4	Decreasing LTPA, L/F§ s1->s4
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0000	206	0.708	0.589	0.622	0.730	0.635	0.563
			0.504, 0.674†	0.559, 0.685	0.637, 0.823	0.554, 0.716	0.452, 0.674
0001	11	0.038	0.033	0.024	0.015	0.014	0.053
			0.008, 0.089	0.004, 0.044	0.003, 0.046	0.001, 0.027	0.010, 0.157
0011	11	0.038	0.046	0.042	0.014	0.025	0.083
			0.015, 0.096	0.010, 0.074	0.003, 0.044	0.007, 0.056	0.027, 0.139
0111	5	0.017	0.046	0.026	0.015	0.026	0.040
			0.004, 0.088	0.007, 0.058	0.006, 0.065	0.005, 0.071	0.008, 0.095
1111	7	0.024	0.071	0.060	0.026	0.049	0.075
			0.032, 0.110	0.025, 0.099	0.006, 0.065	0.007, 0.091	0.039, 0.111

* Proportions of transitions based on frequencies of occurrence in the observed data. The "observed" category, in fact, is one of the set of all possible counterfactual exposure histories. Proportions of transitions, under other counterfactual exposure histories, computed based on estimates from MSM models.

† Confidence intervals that overlap across different exposure histories for given transitions indicate no statistical difference at alpha 0.05 level. Confidence intervals derived from the standard error of a distribution of 1000 bootstrap estimates for each transition probability.

‡ Starting with lowest LTPA category at baseline, there is a 1 unit increase in LTPA category with each successive survey. In conjunction with this increase in LTPA is a lag increase L/F over the population median level (+0.2, +0.5) at time=3, 4.

§ Beginning with the highest LTPA category at baseline, there is a 1 unit decrease in LTPA category with each successive survey. There is a lag decrease in L/F below the population median (-0.2,-0.5) at time=3, 4.

Table 4. Proportions Summed for Selected Groups of Transitions* Over Surveys 1-4 (s1->s4) in Women and Men.

Transitions of Functional Limitation (Yes=1,No=0)	N	Observed Data S1,…S4	Selected Counterfactual Exposure Histories of Lean:Fat Ratio (L/F) and Physical Activity (LTPA)				
			Median LNFAT, No LTPA all surveys	Median LNFAT, High LTPA all surveys	1 Unit over Median LNFAT, No LTPA all surveys	Increasing LTPA, LNFAT s1->s4	Decreasing LTPA, LNFAT s1->s4
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Onset of Functional Recovery (1->0) Transitions†							
Females	**18**	**0.048**	**0.032**	**0.069**	**0.028**	**0.075**	**0.019**
			0.001, 0.063 †	**0.036, 0.102**	**0.008, 0.078**	**0.025, 0.179**	**0.006, 0.049**
Males	**10**	**0.034**	**0.048**	**0.071**	**0.055**	**0.082**	**0.044**
			0.013, 0.111	**0.024, 0.118**	**0.010, 0.154**	**0.018,0.146**	**0.004, 0.084**
Temporal Functional Limitation Transitions‡							
Females	**53**	**0.139**	**0.072**	**0.147**	**0.063**	**0.121**	**0.076**
			0.016, 0.128	**0.093, 0.201**	**0.016, 0.152**	**0.046, 0.196**	**0.027, 0.167**
Males	**34**	**0.117**	**0.118**	**0.139**	**0.114**	**0.165**	**0.083**
			0.042, 0.194	**0.088,0.190**	**0.027, 0.201**	**0.089,0.241**	**0.009, 0.157**
Temporal Functional Recovery Transitions§							
Females Only	**22**	**0.058**	**0.132**	**0.076**	**0.073**	**0.104**	**0.120**
			0.071, 0.193	**0.042, 0.110**	**0.026, 0.152**	**0.028, 0.180**	**0.057, 0.183**

*Particular transitions were grouped and their respective estimated proportions were summed given the characteristics of the transitions and/or infrequency with which they occurred in the sample. Examples of different types of transitions include: † 1->1->1->0, 1->1->0->0, and 1->0->0->0 ; †† 0->0->1->0, 0->1->0->0, 1->0->1->0, and 0->1->1->0; and § 1->0->0->1, 0->1->0->1, 1->0->1->1, and 1->1->0->1.

Discussion

We characterized disablement in a population of elderly subjects as different transitions in functional change over time. Based on estimated MSMs, we estimated the differences in the occurrence of these transitions if the study population sustained different patterns of L/F and LTPA than those it actually experienced. Estimation of differences in functioning under these conditions based on standard statistical methods would not have been possible.

The results suggest that L/F and LTPA affect functioning differently. L/F, which varied minimally with time, appeared to establish levels (strata) of functioning at an early stage of the disablement process that continued over time. For example, we observed that if the population's L/F was one unit greater, the result was a smaller proportion of individuals who experienced limitation at all four surveys and a larger proportion of individuals who were limitation-free over the same period. By contrast, high, sustained levels of LTPA did not increase the proportion of individuals without limitation nor reduce the proportion with a limitation at all four surveys. In fact, in the latter case, the proportion of subjects who were functionally limited at all four surveys increased. However, we observed that high levels of LTPA reduced the risk of onset of functional limitation in subjects without past limitation, and increased the probability of recovery in functioning for those who were previously limited. Based on these findings, we would conclude LTPA reduces the risk of future functional limitation conditional on the level of functioning conferred by L/F.

The data also suggest that the beneficial effects of LTPA with respect to functioning occur indirectly through increase of L/F (i.e. reduction in the amount of fat relative to lean mass). We did not investigate the potential role of past physical activity history on baseline levels of L/F. There is also suggestion of a direct effect of LTPA possibly on some component of functioning (i.e. improved mobility/dexterity). For example, in women, higher levels of physical activity, even without increase in L/F, reduced the onset of functional limitation (table 2, column 4 versus column 7). A comparison of the effects in the men suggested that the advantages conferred by LTPA occur indirectly through L/F. The results were consistent with a previous analysis that indicated LTPA exerts its beneficial effects through reductions in fat mass relative to lean body mass. (5).

One potential limitation of the study could have arisen from failure to satisfy the assumptions that are required to obtain unbiased MSM estimates. For example, we may not have controlled for all measurable confounders of LTPA and L/F and/or misspecified the treatment and censoring models that were used. We have approximated these assumptions by a thorough examination of the potential confounders in our data, and development of treatment/censoring models to control for the effects of confounding and selection bias. We assume we have met all other assumptions that are required to implement the MSM.

We have chosen to examine the causal effect of L/F with respect to functioning. However, to investigate causal effects, the counterfactual outcomes under different levels of the exposure variable must be well-defined. Different processes could have given rise to the same L/F (e.g. exercise, diet), and these different processes could have different implications for the outcome that corresponds to the given L/F. For our purposes, we define L/F as a summary endpoint of these different processes, which in itself has ramifications for functioning regardless of the processes that lead to L/F. Evaluation of the differential effects of L/F on functioning with respect to these different processes was possible , but was not a goal of this analysis.

In summary, our data provide population-level estimates of the extent to which functional limitation can vary over time in the elderly and the potential causal roles of physical activity and body composition in this variation. These causal estimates go beyond what can be inferred from studies that have used more conventional methods of analysis. In addition, our observations point to the need to account for this temporal variation in functioning in the evaluation of any interventions designed to improve functional status in the elderly. Failure to consider this inherent variability could lead to the overestimation of the impacts of interventions and their likely public health significance.

References

1. Anderson RT, James MK, Miller ME, Worley AS, Longino CF, Jr. The timing of change: patterns in transitions in functional status among elderly persons. J Gerontol B Psychol Sci Soc Sci 1998;53:S17-27.
2. Verbrugge LM, Jette AM. The disablement process. Social Science and Medicine 1994;38:1-14.
3. Jette AM. Disability Trends and Transitions. In: Binstock RH, George LK, Marshall VW, Myers GC, Schulz JH, eds. Handbook of Aging and the Social Sciences. New York: Academic Press, 1996:pages 94-116.
4. Sternfeld B, Ngo L, Satariano WA, Tager IB. Associations of body composition with physical performance and self-reported functional limitation in elderly men and women. American Journal of Epidemiology 2002;156:110-21.
5. Tager IB, Haight TJ, Yu Z, Sternfeld B, van der Laan MJ. Longitudinal study of the effects of physical activity and body composition on functional limitation in the elderly. Epidemiology 2004;15:479-493.
6. Janssen I, Heymsfield SB, Ross R. Low relative skeletal muscle mass (sarcopenia) in older persons is associated with functional impairment and physical disability. Journal of the American Geriatrics Society 2002;50:889-96.
7. Stewart AL. Conceptual Challenges in Linking Physical Activity and Disability Research. American Journal of Preventive Medicine 2003;25:137-140.
8. Wolinsky FD, Ambrecht ES, Wyrwich KW. Rethinking functional limitation pathways. The Gerontologist 2000;40:137-46.
9. Jette AM. Assessing disability in studies on physical activity. Am J Prev Med 2003;25:122-8.
10. Seeman TE, Berkman LF, Charpentier PA, Blazer DG, Albert MA, Tinetti ME. Behavioral and psychosocial predictors of physical performance: MacArthur Studies of Successful Aging. Journal of Gerontology: Medical Sciences 1995;50A:M177-83.
11. Miller ME, Rejeski WJ, Beboussin BA, Ten Have TR, Ettinger WH. Physical activity, functional limitation, and disability in older adults. Journal of the American Geriatrics Society 2000;48:1264-72.
12. Berkman LF, Seeman TE, Albert M, et al. High, usual and impaired functioning in community-dwelling older men and women: findings from the MacArthur Foundation Research Network on Successful Aging. Journal of Clinical Epidemiology 1993;46:1129-40.

13. DiPietro L. The epidemiology of physical activity and physical function in older people. Medicine and Science in Sports and Exercise 1996;28:596-600.
14. Stevens J, Cai J, Evenson KR, Thomas R. Fitness and fatness as predictors of mortality from all causes and from cardiovascular disease in men and women in the Lipid Research Clinics Study. American Journal of Epidemiology 2002;156:832-41.
15. Beckett LA, Brock DB, Lemke JH, et al. Analysis of change in self-reported physical function among older persons in four population studies. American Journal of Epidemiology 1996;143:766-78.
16. Gill TM, Kurland BF. Prognostic effect of prior disability episodes among nondisabled community-living older persons. Am J Epidemiol 2003;158:1090-6.
17. Gill TM, Kurland B. The burden and patterns of disability in activities of daily living among community-living older persons. J Gerontol A Biol Sci Med Sci 2003;58:70-5.
18. Hardy SE, Gill TM. Recovery From Disability Among Community-Dwelling Older Persons. JAMA 2004;291:1596-1602.
19. Hardy SE, Dubin JA, Holford TR, Gill TM. Transitions between States of Disability and Independence among Older Persons. American Journal of Epidemiology 2005;161:575-584.
20. Robins JM. Association, causation, and marginal structural models. Synthese 1999;121:151-79.
21. Haight T, Neugebauer R, Tager I, van der Laan M. Comparison of Inverse Probability of Treatment Weighted Estimator and a Naive Estimator. Berkeley Electronic Press/UCB Biostatistics Division 2003;140.
22. Satariano WA, Smith J, Swanson A, Tager IB. A census-based design for the recruitment of a community sample of older residents: Efficacy and costs. Annals of Epidemiology 1998;8:278-82.
23. Tager IB, Hollenberg M, Satariano WA. Association between self-reported leisure-time physical activity and measures of cardiorespiratory fitness in an elderly population. American Journal of Epidemiology 1998;147:921-31.
24. Hollenberg M, Ngo LH, Turner D, Tager IB. Treadmill exercise testing in an epidemiologic study of aging and physical activity: Application of the Cornell Protocol. Journal of Gerontology Biological Sciences 1998;53B:B259-B267.
25. Jette AM, Branch LG. The Framingham Disability Study: Physical disability among the elderly. American Journal of Public Health 1981;71:1211-1216.
26. Cornoni-Huntley J, Brock DB, Ostfeld A. Established Populations for Epidemiologic Studies of the Elderly, Resource Data Book. Bethesda: NIH Publication Number 86-2443, 1986.
27. Nagi SZ. An epidemiology of disability among adults in the United States. Millbank Quarterly 1976;54:439-68.
28. Rosow I, Breslau N. A Guttman health scale for the aged. Journal of Gerontology 1966;21:556-59.
29. Ainsworth BE, Haskell WL, Leon AS, et al. Compendium of physical activities: Classification of energy costs of human physical activities. Medicine and Science in Sports and Exercise 1993;25:71-80.
30. Pate RR, Pratt M, Blair SN, et al. Physical activity and public health: A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. Journal of the American Medical Association 1995;273:402-7.

31. Ferraro KF, Su Y-P, Gretebeck RJ, Black DR, Badylak SF. Body mass index and disability in adulthood: A 20-year panel study. American Journal of Public Health 2002;92:834-40.
32. Flegal KM, Carroll MD, Ogden CL, Johnson CL. Prevalence and trends in obesity among US adults, 1999-2000. Journal of the American Medical Association 2002;288:1723-7.
33. Radloff L. The CES-D scale: a self-report depression scale for research in the general population. Annals of Epidemiology 1977;1:385-401.
34. Satariano WA, Haight TJ, Tager I. Living Arrangements and Participation in Leisure-Time Physical Activities in an Older Population. Journal of Aging and Health 2002;14:427-451.
35. Ross SM. Introduction to probability models. New York: Academic Press, 1972.
36. Hernan M, Brumback B, Robins J. Marginal Structural Models to Estimate Causal Effect of Zidovudine on the Survival of HIV-Positive Men. Epidemiology 2000;11:561-570.
37. Cook NR, Cole SR, Hennekens CH. Use of a marginal structural model to determine the effect of aspirin on cardiovascular mortality in the Physician's Health Study. American Journal of Epidemiology 2002;155:1045-53.
38. Bodnar LM, Davidian M, Siega-Riz AM, Tsiatis AA. Marginal structural models for analyzing causal effects of time-dependent treatments: an application in perinatal epidemiology. Am J Epidemiol 2004;159:926-34.
39. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550-60.
40. Efron B, Tibshirani RJ. An Introduction to the Bootstrap: Chapman and Hall, 1993.

Chapter 4

Secondhand Smoke, Vascular Disease and Dementia Incidence:

Findings From the Cardiovascular Health Cognition Study

Introduction

One of the challenges of aging research is the quantification of effects in the presence of aging processes that could influence the outcome of interest. Previous exposures or past behavioral patterns could affect these aging processes, and/or act independently of them to affect the outcome. To disentangle these effects can present significant analytical challenge. Past effects that may be potentially important with regard to aging outcomes include different health risks, cumulative physical activity accrued over one's lifetime, past education, and environmental and social factors to which one may have been exposed.

Tobacco smoke contains hundreds of chemicals known to be toxic or carcinogenic, and the concentrations of many of these chemicals are higher in secondhand smoke (SHS) than they are in the smoke that is inhaled by smokers(1, 2). Exposure to SHS is associated with developmental and respiratory problems in children as well as increased risk of lung cancer and coronary heart disease in adults(1, 2). In addition, a recent study found that SHS exposure was associated with greater risk of cognitive impairment in adults (3). However, the association between SHS and dementia has not previously been studied.

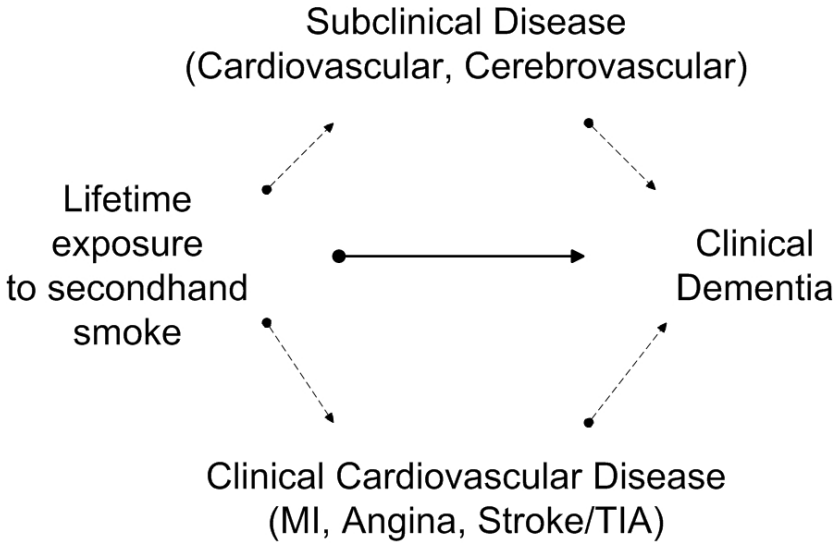
The effects of active smoking on the brain have been somewhat controversial in the past, with some early reports suggesting that active smoking might have beneficial effects (4, 5) or even be associated with a reduced risk of dementia (6, 7). However, more recent evidence suggests that active smoking has neurotoxic effects (8, 9) and is associated with approximately a doubling in dementia risk in older adults (10-12). Therefore, it is plausible that non-smokers who are exposed to high levels of SHS might also experience increased dementia risk.

It also is plausible that SHS might present an indirect risk for dementia through an enhancement of the risks associated with underlying vascular disease. SHS causes a variety of vascular changes including carotid artery thickening, lesion formation, enhanced platelet aggregation, and compromised endothelial function and may contribute to stroke(1, 2, 13, 14). Vascular disease, in turn, has been associated with an increased risk of developing dementia (15). In addition, several recent studies have found that measures of subclinical vascular disease (cerebral magnetic resonance imaging (MRI) findings of small/silent infarcts, enlarged ventricles and white matter disease (16, 17) and ultrasound evidence of carotid artery thickening (18-20)—and their potential co-occurrence with SHS (13, 21) are associated with increased dementia risk and evidence of cognitive impairment.

The primary objective of this study was to determine whether SHS exposure is associated with increased dementia risk among older non-smokers. In addition, given the known deleterious effects of SHS on the vascular system (1, 2), and the growing evidence that vascular disease contributes to the clinical manifestation of dementia (15-19), we hypothesized, *a priori*, that SHS

increases dementia risk in vulnerable subpopulations with underlying clinical or subclinical vascular disease (Figure 1).

Figure 1. Hypothesized Causal Pathways by Which Lifetime Secondhand Smoke (SHS) Exposure Could Increase Risk of Dementia. Secondhand smoke (SHS) could increase risk of dementia directly (solid arrow) or indirectly (dashed arrows) by exacerbating the effects of clinical cardiovascular disease (myocardial infarction [MI], angina, stroke/transient ischemic attack [TIA]) or subclinical cardiovascular or cerebrovascular disease.



Materials and Methods

Subject population

The subject population was participants in the Cardiovascular Health Cognition Study (17, 22, 23), which is nested within the larger Cardiovascular Health Study (CHS) (24). CHS is a prospective, population-based, longitudinal study of risk factors for coronary heart disease and stroke in adults aged 65 years and older. Subjects were recruited from randomized Medicare eligibility lists in four U.S. communities: Forsyth County, NC; Washington County, MD; Sacramento County, CA; and Pittsburgh, PA. CHS enrolled 5,201 participants from 1989-90 and an additional 687 African American participants in 1992-93.

In 1998-99, the CHS Cognition Study was initiated as an ancillary study to identify subjects who had developed dementia during follow-up (17, 22, 23). Participants included 3,608 subjects from both groups who had a cerebral MRI scan and Modified Mini-Mental State (3MS) examination in 1991-94. A standardized protocol was administered across the four sites to classify subjects as having prevalent dementia at the time of the MRI exam or incident dementia from the time of the MRI to the end of the follow-up period (1998-99), death or loss to follow-up.

Of the 3,608 participants, we included only those who had normal cognitive function, were lifelong nonsmokers, and did not have clinical CVD at baseline so that we could focus on

incident disease pathways in participants who were phenotypically healthy. Participants from the African American cohort were excluded from our analyses because it was not possible to differentiate prevalent and incident cardiovascular disease at the time of the MRI in this group; however African American participants from the original cohort were included. Of the 3,171 potential subjects from the original cohort, we excluded 175 subjects who had prevalent dementia at baseline, 415 with mild cognitive impairment for whom year of onset was unknown, 1,335 current or former smokers, 2 subjects with missing smoking status, 244 with underlying cardiovascular disease at baseline, 10 subjects with 3MS scores less than 70 or missing at baseline, and 20 subjects with missing SHS exposure. This left 970 subjects available for analysis.

CHS study procedures were approved by institutional review boards (IRB) at each site, and all participants signed an informed consent at entry and periodically throughout the study. In addition, the secondary data analyses described here were approved by the CHS Steering Committee; the Committee on Human Research at the University of California, San Francisco; the San Francisco Veteran's Administration Medical Center R&D Committee; and the IRB at the University of California, Berkeley.

Dementia diagnosis

Dementia was defined as a progressive or static deficit in at least two cognitive domains that did not necessarily include memory and was of sufficient severity to affect the subjects' daily activities combined with a previous history of normal intellectual function (17, 22, 23). This definition differs slightly from the standard Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, definition(25), which requires a memory deficit. Individuals who did not meet dementia criteria but who exhibited poor cognitive function that reflected a decline from a prior level were classified as having mild cognitive impairment. Standard criteria were used to classify dementia type as probable or possible Alzheimer's Disease, probable or possible vascular dementia, mixed dementia, or other(25-28).

Diagnoses were made based on a review of available data by an adjudication committee that consisted of neurologists and psychiatrists (one from each site) with expertise in dementia diagnosis (17, 22, 23). Data available for review consisted of information collected annually as part of the main CHS study and included cognitive test scores, depressive symptoms, level of difficulty with activities of daily living and instrumental activities of daily living, hearing and vision problems, alcohol intake, use of drugs to treat dementia, and recent hospital records. In addition, more detailed neuropsychological, neurological and neuropsychiatric data were available in a sub-group of high-risk participants who were examined in 1998-99. Subjects who died during follow-up were classified based on their status at the time of death; those who were lost to follow-up were classified based on available data up to their last evaluation.

Secondhand smoke

Subjects were asked whether they had ever lived with anyone who smoked cigarettes regularly and, if so, the total number of years and the time period (childhood, ages 20-50, after age 50). Exposure variables were created to reflect the total number of years of exposure during

each of these time periods. Deciles of risk of SHS exposure were created based on subjects' cumulative years of exposure, and analyses were carried out to examine the risk of dementia given these different levels. Based on preliminary analyses of the distribution of SHS exposure, three categories of SHS exposure were created for further analyses: none/low (≤15 years), moderate (16-25 years), and high (>25 years). Subjects with no exposure and low exposure were combined into the lowest category since there was no observed difference in risk of dementia between these two groups. Workplace SHS exposure was not assessed.

Vascular disease measures

Clinical vascular disease (CVD) was defined as having a history of myocardial infarction (MI), stroke, transient ischemic attack (TIA), angina pectoris, claudication, angioplasty or by-pass surgery. All vascular events were identified at baseline and during follow-up as part of the main CHS study using a rigorous protocol that required validation by either physician questionnaire or medical record review (24, 29, 30).

Subclinical magnetic resonance imaging (MRI) measures were based on cerebral MRI exams performed using a standard protocol (31, 32). Images were interpreted by trained neuroradiologists who were blinded to subjects' age, sex, race, ethnicity and other clinical information. Infarcts on MRI were defined as lesions with abnormal signal in a vascular distribution and no mass effect. White matter disease was estimated as the total volume of periventricular and subcortical white matter signal abnormality on spin density-weighted axial images compared with 8 'reference' images and was classified from grade 0 (none) to 9 (extensive). Specific subclinical MRI measures used in this study included small infarcts (< 3mm), large infarcts (≥ 3mm) and white matter disease (grade 3 or more).

Subclinical carotid artery measures were based on Duplex ultrasonography performed with two-dimensional brightness mode imaging to detect thickening of the arterial wall, disruption of normal wall surfaces and development of focal plaques bilaterally.(33) Images were interpreted at the CHS Ultrasound Reading Center by trained readers. Specific subclinical carotid artery measures used in this study included internal or common carotid artery thickness above the 80^{th} percentile and stenosis > 25% of the internal carotid artery.

Other measures

Time-independent measures included age (categorized as <70, 70-73, 74-79, 80+ years), race (African-American/Other, Caucasian), gender, income (categories for range of <$5,000->$50,000), education (high school equivalent or greater), apolipoprotein-E genotype (ε4 allele present/absent), C-reactive protein (ml/L), and occupation (professional, sales/clerical, farmer/craftsman, housewife, other). Time-dependent measures included self-reported health (excellent, good, fair, or poor), hypertension (ever hypertensive/borderline or not hypertensive based on history of hypertension and/or measured blood pressure at visit), diabetes (American Diabetes Association criteria: fasting glucose <126 mg/dl ≥126 mg/dl and/or oral hypoglycemic/insulin therapy), physical activity (blocks walked in the last week), depression

(Center for Epidemiologic Studies-Depression (34) Scale score of 16 or greater), weight (kg), cholesterol (mg/dl) and alcohol use (number of beverages per week). The time-dependent measures were recorded annually over follow-up with the exception of the diabetes measure which was recorded for 2 of the possible 6 annual visits. These variables were accounted for as potential confounders in the analysis (see Appendix 3).

Statistical analysis

Bivariate and multiway associations between subclinical and clinical CVD, SHS exposure and dementia were examined. Associations between potential confounding factors and dementia (e.g., age) also were evaluated. Stratified Kaplan-Meier plots were used to examine the distribution of incident dementia by subclinical and clinical CVD status.

Cox proportional hazards marginal structural models (Cox PH MSMs) were used to study different causal pathways between SHS and dementia risk (Figure 1). Estimation of these models accounted for clinical CVD on the causal pathway. The estimation procedures that were applied can provide unbiased estimates of causal effects in the context of time-dependent confounders and causal intermediates (35-38), whereas standard analytic methods are likely to produce biased risk estimates under these conditions (39-41). Details of the application of MSM methodology in this study are provided in the Appendix 3 and elsewhere (39).

For our analyses, Cox PH MSMs were fit using weighted logistic regression estimated with generalized estimating equations, with individual weights derived for each subject (see Appendix 3) (36). We constructed a series of Cox PH MSMs to examine different mechanisms by which SHS exposure might affect risk of incident dementia. These included the direct effects of SHS and CVD (Model 1); the additional direct effects of subclinical MRI measures (Model 2) and subclinical carotid artery measures (Model 3); and the joint effects of SHS and subclinical carotid artery measures (Model 4). For some of the models that were posited (e.g., joint effects of SHS and CVD, joint effects of SHS and MRI measures), sample sizes were too small to test our hypotheses. All models accounted for CVD as a causal intermediate; therefore, hazard ratio (HR) estimates from these models reflect associations that are independent of any effects through CVD.

All analyses were performed with SAS version 9.1.3.

Results

Descriptive and unadjusted analyses

Subjects had a mean (standard deviation [SD]) age of 74 (5) years and 74% were women. Over 60% had lived with a smoker for ≤15 years (N=600), including 470 who had never lived with a smoker; 13% (N=130) had lived with a smoker for 16-25 years, and 25% (N=247) had lived with a smoker for >25 years. Subjects were followed for a mean of 5.5 years (range: 0.5 – 8.4), during which time 15% (N=148) developed dementia (94 Alzheimer's disease, 41 vascular dementia, 10 mixed dementia, 3 other).

Participants who had lived with a smoker for ≤15 years were older and were less likely to be female compared to those who had lived with a smoker for more than 15 years (Table 1). They also were more likely to have common carotid artery thickness above the 80th percentile. However, there were no differences between the three SHS groups based on education or

presence of clinical vascular disease, subclinical MRI measures or subclinical carotid artery measures.

In unadjusted analyses, there was no evidence of an association between SHS exposure and risk of dementia (Table 2). In addition, clinical vascular disease was not significantly associated with dementia risk. In contrast, most of our measures of subclinical vascular disease were associated with increased dementia risk. Specifically, risk of dementia was increased in participants with large infarcts, small infarcts or white matter disease on the MRI scans as well as those with internal or common carotid artery thickness above the 80th percentile.

MSM analyses

In our MSM analyses (Table 3), clinical CVD was the 'causal effects' parameter while the other variables in the model were treated as 'stratification' variables (see Appendix). Therefore, the hazard ratio estimates for clinical CVD reflect the population-level change in the relative hazard of incident dementia if, contrary to fact, everyone in the population experienced clinical CVD compared to if no one experienced clinical CVD. In contrast, the hazard ratio estimates for the other variables in the models reflect the change in the relative hazard of dementia associated with a particular exposure if, contrary to fact, no one in the population experienced clinical CVD (i.e., independent of any effects of clinical CVD).

Model 1 indicates that the population-level relative hazard of dementia was estimated to be increased by 65% in those with CVD if, contrary to fact, no one in the population experienced CVD, although this increase was not statistically significant (HR 1.65; 95% CI: 0.62, 3.16). Additionally if, contrary to fact, no one experienced clinical CVD in the population, there was no evidence of an associated change in the relative hazard of dementia in those with moderate SHS exposure (HR, 1.02; 95% CI: 0.48, 1.88) or high SHS exposure (HR, 1.43; 95% CI: 0.80, 2.32) relative to those with low/no SHS exposure.

When subclinical MRI measures and subclinical carotid artery measures were added, there was little change in the relationships between clinical CVD, SHS exposure and dementia risk. However, if, contrary to fact, no one in the population experienced clinical CVD, then the associated relative hazard of dementia was estimated to increase by two-thirds in those with small MRI infarcts (HR, 1.67; 95% CI: 0.92, 2.98) and was more than 2.5 times higher in those with evidence of white matter disease (HR, 2.65; 95% CI: 1.70, 4.34). The associated relative change in hazard of dementia was not significantly altered by large MRI infarcts (Model 2) or subclinical carotid artery measures (Model 3) after removing the effects of clinical CVD.

When the joint effects of SHS and subclinical vascular measures were examined (Model 4), there was evidence of interaction between SHS exposure and internal carotid artery stenosis on dementia risk. If, contrary to fact, no one experienced clinical CVD, neither SHS exposure nor internal carotid artery stenosis alone were associated with dementia risk (Figure 2; Table 3, Model 4); however, the relative hazard of dementia for those with both >25% stenosis and >25

Table 1. Baseline Characteristics of 970 Non-Smokers by Level of Second-Hand Smoke (SHS) Exposure, Cardiovascular Health Cognition Study, 1991-1994

	Level of SHS Exposure			
Variable	**0-15 years (N=600)**	**16-25 years (N=123)**	**>25 years (N=247)**	***P*-value**
Demographics				
Age, yr	75.0 (4.9)	73.9 (4.3)	73.8 (4.3)	<0.001
Women (%)	68.7	82.1	83.0	<0.001
Education (% ≥ high school diploma)	76.6	84.3	78.7	0.12
Clinical vascular disease (%)	11.3	9.8	9.7	0.74
Subclinical MRI measures				
MRI large infarcts (%)	28.8	22.8	23.1	0.14
MRI small infarcts (%)	13.6	16.3	11.7	0.48
MRI white matter disease (%)	32.1	24.6	32.7	0.23
Subclinical carotid artery measures				
Internal artery thickness > 80^{th} percentile (%)	20.2	18.7	19.6	0.92
Common artery thickness > 80^{th} percentile (%)	21.8	13.8	15.9	0.04
Internal artery stenosis > 25%	39.0	35.0	34.7	0.42

Values are mean (standard deviation) or %. *P*-values based on the F-test for continuous data and the χ^2 test for categorical data.

Table 2. Unadjusted Association between Second-Hand Smoke, Vascular Measures and Dementia Incidence.*

Characteristic	Dementia N (%)†	No Dementia N (%)	CIR (95% CI)‡
Secondhand Smoke			
0-15 years	95 (15.8)	505 (84.2)	Ref
16-25 years	17 (13.8)	106 (86.2)	0.87 (0.54, 1.41)
>25 years	36 (14.6)	211 (85.4)	0.92 (0.65, 1.31)
Clinical vascular disease			
Absent	138 (15.9)	728 (84.1)	Ref
Present	10 (9.6)	94 (90.4)	0.60 (0.33, 1.11)
Subclinical MRI measures			
Large infarcts $\geq$ 3mm			
Absent	94 (13.2)	616 (86.8)	Ref
Present	52 (20.2)	205 (79.8)	1.53 (1.12, 2.08)
Small infarcts < 3mm			
Absent	117 (14.0)	720 (86.0)	Ref
Present	29 (22.3)	101 (77.7)	1.60 (1.11, 2.29)
White matter disease			
Absent	64 (9.7)	595 (90.3)	Ref
Present	81 (27.0)	219 (73.0)	2.78 (2.06, 3.74)
Subclinical carotid artery measures			
Internal artery thickness > 80%			
Absent	103 (13.4)	667 (86.6)	Ref

Present	42 (22.0)	149 (78.0)	1.64 (1.19, 2.27)
Common artery thickness > 80%			
Absent	105 (13.5)	671 (86.5)	Ref
Present	40 (21.6)	145 (78.4)	1.60 (1.15, 2.22)
Stenosis ≥ 25%			
Absent	83 (13.8)	518 (86.2)	Ref
Present	62 (17.3)	297 (82.7)	1.25 (0.92, 1.69)

CI, confidence interval; CIR, cumulative incidence ratio; MRI, magnetic resonance imaging.
*Missing values between 0-10 for different characteristics.
†148 cases of dementia included 94 subjects with Alzheimer's disease only; 41 subjects with vascular dementia; 10 subjects with mixed vascular dementia/Alzheimer's Disease; and 3 subjects with other types of dementia.
‡Unadjusted CIR calculated based on dementia status at last follow-up visit.

Table 3. Effects of Secondhand Smoke and Vascular Disease on Dementia Incidence Using Cox Proportional Hazards Marginal Structural Models.

	Hazard Ratio (95% Confidence Interval)*			
Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Clinical vascular disease†	1.65 (0.62, 3.16)	1.56 (0.53, 3.10)	1.59 (0.62, 3.21)	1.60 (0.60, 3.27)
Secondhand smoke (SHS)‡				
16-25 years	1.02 (0.48, 1.88)	1.08 (0.53, 2.02)	1.02 (0.48, 1.90)	1.13 (0.38, 2.43)
>25 years	1.43 (0.80, 2.32)	1.28 (0.71, 2.14)	1.46 (0.82, 2.40)	0.81 (0.34, 1.50)
Subclinical MRI Measures§				
Infarct ≥ 3mm		0.90 (0.50, 1.62)		
Infarct < 3mm		1.67 (0.92, 2.98)		
White Matter Disease (WMD)		2.65 (1.70, 4.34)		
SHS 16-25 yrs*WMD				
SHS >25 yrs*WMD				
Subclinical Carotid Artery Measures§				
Stenosis >25%			0.99 (0.59, 1.67)	0.76 (0.40, 1.41)
Common wall thickness > 80th percentile			1.07 (0.66, 1.69)	1.08 (0.66, 1.73)

Internal wall thickness >80th percentile	1.60 (0.80, 3.07)	1.52 (0.77, 2.92)
SHS 16-25 yrs*Stenosis		0.87 (0.21, 3.35)
SHS >25 yrs* Stenosis		3.00 (1.03, 9.72)

MRI, magnetic resonance imaging; SHS, secondhand smoke; WMD, white matter disease

*All models adjusted for age, gender, and education. Other variables listed in the Methods section were included in the 'treatment' model and, therefore, it was not necessary to adjust for them in the MSM models. See Appendix for more details.

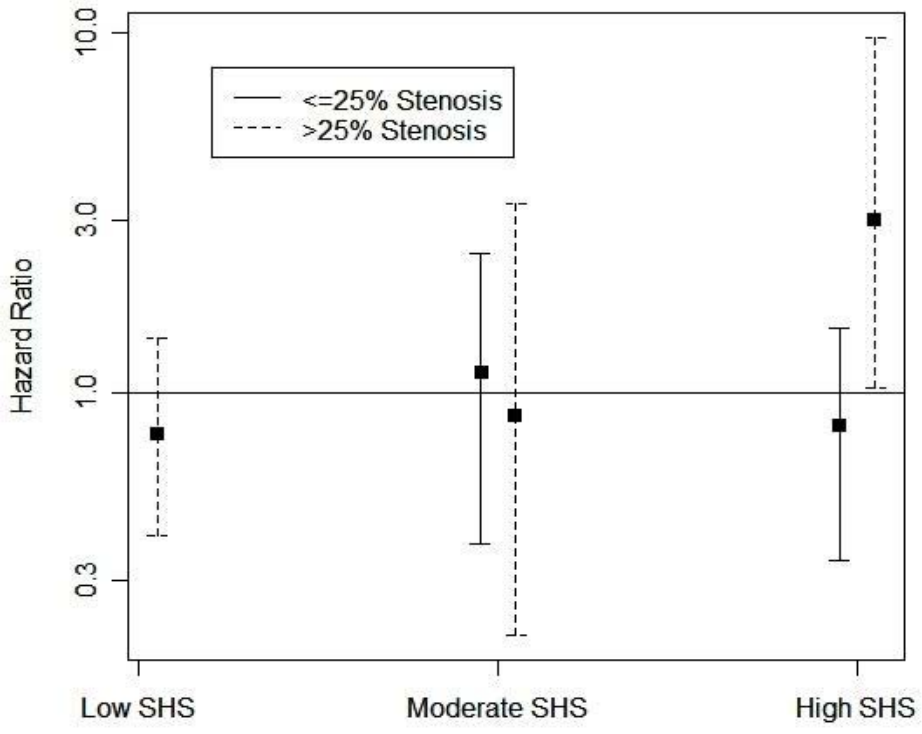
†Marginal (population) relative hazard of clinical cardiovascular disease (CVD) on dementia at *t* for a given stratum of age, secondhand smoke (SHS) exposure, gender and education.

‡Associated relative hazard of dementia at *t* for various SHS exposure levels compared to those with 0-15 years of SHS exposure for a given stratum of age, gender, and education, independent of any effects of CVD.

§Associated relative hazard of dementia at *t* in subjects for various subclinical MRI measures (Model 2), subclinical carotid artery measures (Model 3) or interactions between SHS and subclinical MRI measures (Model 4) for a given stratum of age, gender, education, and SHS exposure, independent of any effects of CVD.

Results were similar when subclinical MRI measures and subclinical carotid artery measures were included in the same model to determine whether their effects were independent, and when Modified Mini-Mental State Exam score was included as stratification variable to further control for residual confounding by baseline cognitive function, education and socioeconomic status.

FIGURE 2. Secondhand Smoke (SHS), Internal Carotid Artery Stenosis and Dementia Risk. Cox proportional hazards marginal structural models were used to calculate hazard ratios for dementia as a function of secondhand smoke (SHS) exposure and internal carotid artery stenosis. Neither SHS nor internal carotid artery stenosis increased dementia risk when considered alone. However, the risk of dementia was three times higher in those with high levels of SHS exposure (>25 years) and internal carotid artery stenosis >25% compared to those with no/low SHS exposure (<15 years) and ≤25% stenosis.



years of SHS exposure was three times higher (HR, 3.00; 95% CI: 1.03, 9.72) than those with neither of these characteristics. There was no suggestion that the relative hazard of dementia associated with other subclinical cardiovascular disease variables increased as the level of SHS exposure increased (data not shown).

Discussion

In this study of almost 1,000 lifetime non-smokers, we found that exposure to high levels of SHS in combination with carotid artery stenosis was associated with an elevated risk of developing dementia over six years. The risk of estimated incident dementia was tripled in participants who had lived with a smoker for more than 25 years over their lifetimes and also had carotid artery stenosis > 25% at baseline. There was no evidence of a direct effect of SHS exposure on the risk of incident dementia independent of the pathway through carotid artery stenosis.

This is the first study to examine the potential causal association of lifetime exposure to SHS and dementia in older adults. However, its findings are consistent with studies that have found an association between SHS exposure and worse cognitive function in children (8, 42, 43) and, more recently, adults (3). A study of more than 4,000 children age 6 to 16 years who participated in the Third National Health and Nutrition Survey found an inverse relationship between serum cotinine—a marker of tobacco smoke exposure—and performance on tests of reading, math and block design even after controlling for sex, race, region, poverty, parent education and marital status, ferritin and blood lead concentration. Similarly, a national population-based study of more than 4,000 adults in England found that non-smokers with high SHS exposure based on cotinine levels had greater odds of cognitive impairment (3).

Several factors support the biological plausibility of SHS exposure as a risk factor for dementia. First, SHS is highly toxic and contains at least 250 chemicals known to be harmful or carginogenic (2); therefore, such exposure could negatively impact the brain and render it more susceptible to dementia. Second, exposure to SHS can cause both immediate and long-term adverse effects in the cardiovascular system that include increased ‘stickiness’ of blood platelets, damage to the lining of blood vessels, endothelial dysfunction, decreased coronary flow velocity reserves and reduced heart rate variability (2, 13). Studies have found that endothelial dysfunction may be related to the reduced clearance of beta-amyloid protein, which is considered to be one of the factors related to the pathogenesis of Alzheimer’s disease (44). Third, there is evidence that nonsmokers exposed to high levels of SHS may develop atherosclerosis of the carotid and large arteries of the brain as well as degeneration of the intracerebral arteries. These changes, in turn, may increase the risk of stroke and dementia (1, 13, 15, 21). Therefore, there are several plausible mechanisms through which SHS could exert direct and/or indirect effects on the risk of dementia.

In this study, the hazard of dementia was increased three-fold in subjects with >25 years SHS exposure and >25% stenosis of the internal carotid artery, although the estimate of the interaction was imprecise (95% CI: 1.03, 9.72). There was no evidence of an associated increased risk of dementia due to SHS exposure or carotid artery stenosis alone. It is possible that the effects of these variables were not sufficient to represent independent risk factors, but their combined effects may have acted synergistically to induce the observed association with dementia. Similar interactions between SHS

exposure and other measures of underlying cardiovascular disease (e.g., MRI measures, intimal wall thickness) were not observed, however. Absence of other interactions was partly due to insufficient data. However, our ability to observe other interactions may have been limited for biological reasons as well. The fact that a weak, non-significant interaction (data not shown) was observed between categories of SHS exposure and intimal wall thickness IMT (>80th percentile of the internal carotid artery) may have been due to insufficient data (i.e., imprecise estimates), as well as the fact that the study population consisted mainly of healthy older women. The distribution of IMT in these individuals, given it was a relative measure, may not have been of sufficient severity to observe either an independent effect of IMT or a joint effect between IMT and SHS exposure. The absence of interactions between SHS and the MRI measures (i.e., WMD, large infarcts), and associated dementia risk, might be explained by the fact that the MRI variables are measures of underlying disease in the brain itself, and may possibly represent distinct pathways, separate from those that involve general underlying CVD, to dementia risk. It is plausible that the added effects of SHS exposure by way of these pathways might be negligible when compared with the risk of positive levels of different MRI variables for dementia and cerebrovascular disease.

Strengths of the study include its well-characterized study population, detailed measures of both clinical and subclinical vascular disease, the specificity of the study design, and the use of causal inference methodology, which made investigation of the different causal pathways between SHS and dementia possible. A limitation was that SHS was based on self-report and did not include work exposure. However, the study population consisted mainly of older women (~74%) whose most likely source of SHS was household exposure.

The study was restricted to nonsmokers who had no pre-existing clinical dementia or cardiovascular disease. Specific types of subclinical CVD that were prevalent in this population were examined in combination with different levels of SHS exposure to represent selective biological pathways by which SHS might differentially affect risk of dementia. However, the additional specificity that was incorporated as part of the study reduced the sample size of the analysis and likely decreased the precision of our estimates of effect.

It is possible that the statistically significant interaction observed between SHS and carotid artery stenosis was due to chance (i.e., Type I error). Correction for multiple comparisons that used a resampling-based widened the confidence interval slightly (95% CI: 0.97, 10.32); however, the point estimate remained unchanged (HR, 3.00) (45, 46). Correction for multiple comparisons has remained controversial in practice, particularly when underlying causal relationships are suspected, because of an increased probability of Type II error (47).

This study uses an analysis method that addresses the fact that clinical cardiovascular disease reflects, for many people, a causal pathway between SHS exposure and cerebrovascular disease. Moreover, we hypothesized that clinical cardiovascular disease represents a causal pathway between SHS exposure and incident dementia. Thus a direct effect of SHS on dementia risk could be examined, as well as the potential contribution of SHS exposure in subjects with subclinical levels of CVD which are known to have associated risk with dementia. Causal inference methods—i.e., a Cox PH MSM and estimation of that model as used in this analysis-- provide a method for

control of confounding factors that are hypothesized to be on the causal pathway (e.g., clinical CVD in this study), which is not possible with standard statistical approaches. The application of MSMs in this instance has relevant public health implications because it provides a quantitative basis for assessment of the need to target older individuals whose underlying characteristics put them at greater risk of dementia.

In summary, in this cohort of non-smoking elders, we found that exposure to high levels of SHS alone did not increase dementia risk. However, those elders who had a history of high SHS exposure combined with a history of carotid artery stenosis experienced a three-fold increase in the risk of dementia.

References

1. California Environmental Protection Agency: Air Resources Board and Office of Environmental Health Hazard Assessment. Proposed identification of environmental tobacco smoke as a toxic air contaminant as approved by the Scientific Review Panel on June 24, 2005 (Appendix III). Sacramento, CA: California Environmental Protection Agency,, 2005.
2. US Department of Health and Human Services. The Health Consequences of Involuntary Exposure to Tobacco Smoke: A Report of the Surgeon General. Rockville, MD: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health;, 2006.
3. Llewellyn DJ, Lang IA, Langa KM, Naughton F, Matthews FE. Exposure to secondhand smoke and cognitive impairment in non-smokers: national cross sectional study with cotinine measurement. Bmj 2009;338:b462.
4. Elrod K, Buccafusco JJ, Jackson WJ. Nicotine enhances delayed matching-to-sample performance by primates. Life Sci 1988;43:277-87.
5. Salomon AR, Marcinowski KJ, Friedland RP, Zagorski MG. Nicotine inhibits amyloid formation by the beta-peptide. Biochemistry 1996;35:13568-78.
6. Lee PN. Smoking and Alzheimer's disease: a review of the epidemiological evidence. Neuroepidemiology 1994;13:131-44.
7. Van Duijn CM, Clayton DG, Chandra V, et al. Interaction between genetic and environmental risk factors for Alzheimer's disease: a reanalysis of case-control studies. EURODEM Risk Factors Research Group. Genet Epidemiol 1994;11:539-51.
8. Swan GE, Lessov-Schlaggar CN. The effects of tobacco smoke and nicotine on cognition and the brain. Neuropsychol Rev 2007;17:259-73.
9. Reitz C, Luchsinger J, Tang MX, Mayeux R. Effect of smoking and time on cognitive function in the elderly without dementia. Neurology 2005;65:870-5.
10. Almeida OP, Hulse GK, Lawrence D, Flicker L. Smoking as a risk factor for Alzheimer's disease: contrasting evidence from a systematic review of case-control and cohort studies. Addiction 2002;97:15-28.
11. Anstey KJ, von Sanden C, Salim A, O'Kearney R. Smoking as a risk factor for dementia and cognitive decline: a meta-analysis of prospective studies. Am J Epidemiol 2007;166:367-78.

12. Hernan MA, Alonso A, Logroscino G. Cigarette smoking and dementia: potential selection bias in the elderly. Epidemiology 2008;19:448-50.
13. Barnoya J, Glantz SA. Cardiovascular effects of secondhand smoke: nearly as large as smoking. Circulation 2005;111:2684-98.
14. Zhang X, Shu XO, Yang G, et al. Association of passive smoking by husbands with prevalence of stroke among Chinese women nonsmokers. Am J Epidemiol 2005;161:213-8.
15. Snowdon DA, Greiner LH, Mortimer JA, Riley KP, Greiner PA, Markesbery WR. Brain infarction and the clinical expression of Alzheimer disease. The Nun Study. Jama 1997;277:813-7.
16. Vermeer SE, Prins ND, den Heijer T, Hofman A, Koudstaal PJ, Breteler MM. Silent brain infarcts and the risk of dementia and cognitive decline. N Engl J Med 2003;348:1215-22.
17. Kuller LH, Lopez OL, Newman A, et al. Risk factors for dementia in the cardiovascular health cognition study. Neuroepidemiology 2003;22:13-22.
18. Mathiesen EB, Waterloo K, Joakimsen O, Bakke SJ, Jacobsen EA, Bonaa KH. Reduced neuropsychological test performance in asymptomatic carotid stenosis: The Tromso Study. Neurology 2004;62:695-701.
19. Kearney-Schwartz A, Rossignol P, Bracard S, et al. Vascular structure and function is correlated to cognitive performance and white matter hyperintensities in older hypertensive patients with subjective memory complaints. Stroke 2009;40:1229-36.
20. Newman AB, Fitzpatrick AL, Lopez OL, et al. Dementia and Alzheimer's Disease incidence in relationship to cardiovascular disease in the Cardiovascular Health Study Cohort. Journal of the American Geriatric Society 2005; 53:1101–1107.
21. Hanon O, Haulon S, Lenoir H, et al. Relationship Between Arterial Stiffness and Cognitive Function in Elderly Subjects With Complaints of Memory Loss. Stroke 2005;36:2193-2197.
22. Fitzpatrick AL, Kuller LH, Ives DG, et al. Incidence and Prevalence of Dementia in the Cardiovascular Health Study. Journal of American Geriatric Society 2004;52:195-204.
23. Lopez OL, Kuller LH, Fitzpatrick A, Ives D, Becker JT, Beauchamp N. Evaluation of dementia in the cardiovascular health cognition study. Neuroepidemiology 2003;22:1-12.
24. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol 1991;1:263-76.
25. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV. Washington D.C.: American Psychiatric Association;, 1994.
26. Chui HC, Victoroff JI, Margolin D, Jagust W, Shankle R, Katzman R. Criteria for the diagnosis of ischemic vascular dementia proposed by the State of California Alzheimer's Disease Diagnostic and Treatment Centers. Neurology 1992;42:473-80.
27. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group

under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984;34:939-44.
28. Roman GC, Tatemichi TK, Erkinjuntti T, et al. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. Neurology 1993;43:250-60.
29. Ives DG, Fitzpatrick AL, Bild DE, et al. Surveillance and ascertainment of cardiovascular events. The Cardiovascular Health Study. Ann Epidemiol 1995;5:278-85.
30. Psaty BM, Kuller LH, Bild D, et al. Methods of assessing prevalent cardiovascular disease in the Cardiovascular Health Study. Ann Epidemiol 1995;5:270-7.
31. Bryan RN, Manolio TA, Schertz LD, et al. A method for using MR to evaluate the effects of cardiovascular disease on the brain: the cardiovascular health study. AJNR Am J Neuroradiol 1994;15:1625-33.
32. Manolio TA, Kronmal RA, Burke GL, et al. Magnetic resonance abnormalities and cardiovascular disease in older adults. The Cardiovascular Health Study. Stroke 1994;25:318-27.
33. O'Leary DH, Polak JF, Kronmal RA, et al. Distribution and correlates of sonographically detected carotid artery disease in the Cardiovascular Health Study. The CHS Collaborative Research Group. Stroke 1992;23:1752-60.
34. Radloff L. A self-report depression scale for research in the general population. Applied Psycological Measure 1977;1:385-401.
35. Eisner M, Wang Y, Haight T, Balmes J, Hammond K, Tager I. Second-hand smoke exposure, pulmonary function, and cardiovascular mortality. Annals of Epidemiology 2007;17:364-373.
36. Hernan M, Brumback B, Robins J. Marginal Structural Models to Estimate Causal Effect of Zidovudine on the Survival of HIV-Positive Men. Epidemiology 2000;11:561-570.
37. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550-60.
38. Tager IB, Haight TJ, Yu Z, Sternfeld B, van der Laan MJ. Longitudinal study of the effects of physical activity and body composition on functional limitation in the elderly. Epidemiology 2004;15:479-493.
39. Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. Epidemiology 2006;17:276-84.
40. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology 1992;3:143-55.
41. Weinberg CR. Toward a clearer definition of confounding. Am J Epidemiol 1993;137:1-8.
42. Rush D, Callahan KR. Exposure to passive cigarette smoking and child development. A critical review. Ann N Y Acad Sci 1989;562:74-100.
43. Yolton K, Dietrich K, Auinger P. Exposure to Environmental Tobacco Smoke and Cognitive Abilities among U.S. Children and Adolescents. Environmental Health Perspectives 2005;113:98-103.
44. Zlokovic BV. Neurovascular mechanisms of Alzheimer’s neurodegeneration. Trends Neurosci 2005;28:202-208.

45. Pollard KS, van der Laan MJ. Choice of a null distribution in resampling-based multiple testing. Journal of Statistical Planning and Inference 2004;125:85-100.
46. van der Laan MJ, Robins J. Unified Methods for Censored Longitudinal Data and Causality, 2002.
47. Rothman K. No Adjustments Are Needed for Multiple Comparisons. Epidemiology 1990;1:43-46.

Chapter 5

Direct effects of leisure-time physical activity on walking speed

Introduction

While long-term effects of LTPA may contribute to patterns of physical function over time, as was shown in an earlier chapter, one may be interested, too, in its short-term effects on outcomes that it may influence as the aging process unfolds. Walking speed is an example of such an outcome that declines with age, but at different rates among individuals, and is used by clinicians' as an indicator of their patients' future functional status (1-4). Therefore, it behooves researchers to consider some of the mechanisms by which it could be maintained in an older population.

It is well known that leisure-time physical activity (LTPA) is associated with several important health benefits across all age-groups (5-7). Individuals who participate in LTPA have been shown to be at a reduced risk of hypertension, colon cancer, diabetes, cardiovascular disease, and premature mortality (5, 6, 8). Moreover, high intensity LTPA increases cardiorespiratory fitness, and moderate levels have been associated with improved muscle strength, and metabolic function (e.g., glucose tolerance, lipid utilization) (5, 9-11) In the elderly, LTPA is associated with reduced risk of cardiovascular disease, increased physical and cognitive functioning, and reduced risk of falls and disability(7, 12).

Walking speed is a measure of lower body strength and function. It is also an important predictor of future disability and mortality (1-4, 13, 14). It is likely that LTPA preserves walking speed through a variety of pathways (e.g., muscoskeletal, neurologic, cardiovascular) (15-18). For example, it is likely that LTPA preserves walking speed in the elderly partly through maintenance of a lower body mass index (BMI) and preservation of lean mass relative to fat mass (9, 15, 16, 19). Studies have shown a positive association between higher lean mass, lower fat mass and increased walking speed and improved lower body function in elderly subjects (20, 21). Other studies have found that fat mass and relative measures of body composition (e.g., lean mass relative to fat mass) may be more predictive of walking speed than lean mass alone (22, 23).

While LTPA is likely to contribute to the physiologic systems that are responsible for increased walking speed (improved aerobic capacity, muscle strength), little is known about the relative contributions of LTPA to each of these various pathways. Examination of the contributions of LTPA to these various mechanisms is important when one considers the increased prevalence of diseases such as obesity and diabetes in older adults as well as the general public (24). Given these older populations, these diseases and other age-related changes may act in conjunction to reduce the number of possible pathways by which LTPA may benefit walking speed (24, 25).

One of the mechanisms through which LTPA is thought to exert its health benefits is the maintenance of glycemic control and prevention of insulin resistance (25,

26). Insulin resistance is typically associated with lower glycemic control (i.e., higher fasting glucose), higher percentage fat mass compared to lean mass, and is a risk factor for type 2 diabetes (T2D) (9, 27). Exercise is typically prescribed to individuals with insulin resistance and T2D to increase insulin sensitivity, regain glycemic control through glucose uptake, and restore individuals' energy balance (9, 28). However, it is uncertain whether exercise has an equal benefit in the elderly with respect to glucose uptake and increased insulin sensitivity given the reduced capacity of various cellular respiratory pathways (i.e., reduced activity of oxidative enzymes, defective glucose transportation) that can occur with aging(29-31) If so, older adults would be less likely to benefit from effects of LTPA on walking speed through higher lean mass relative to fat mass than their younger counterparts. Alternatively, this group may still benefit from LTPA through alternative pathways (e.g., neurologic, cardiovascular-- such as reduced peripheral artery disease and inflammation, increased blood flow) not mediated by body composition (Figure 1).

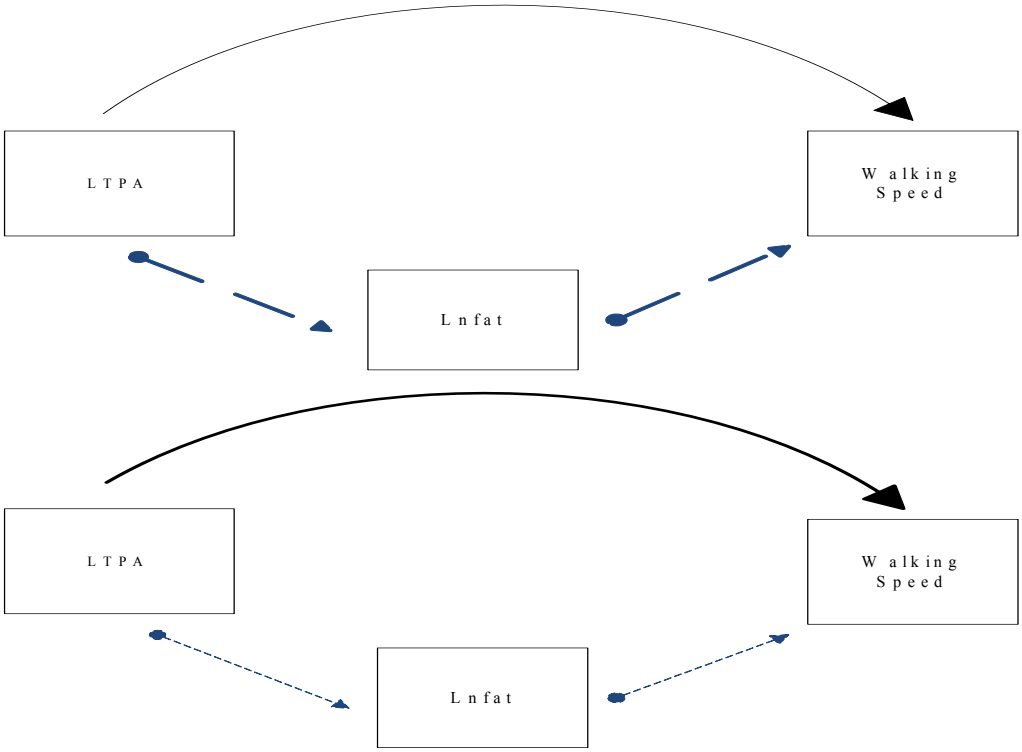


Figure 1. Hypothesized direct and indirect pathways from leisure-time physical activity (LTPA) to walking speed with lean:fat mass (Lnfat) as a causal intermediate in the elderly. The upper graph(darkened arrows) reflect the hypothesis that most of the effects of LTPA on walking speed are mediated through Lnfat in healthy individuals. The lower graph (darkened curvilinear line) reflects the hypothesis that relatively greater direct effects of LTPA (through combination of alternative pathways) occur in those with

metabolic derangements (e.g., insulin resistance, glucose intolerance) related to older age and/or diabetes and cardiovascular disease.

The aim of the current study is to investigate the short-term direct effects of LTPA (i.e., physical activity reported in the previous year) on walking speed through pathways other than those to body composition as measure by lean:fat mass ratio (Lnfat). In this context, Lnfat represents surrogate for the metabolic derangements discussed above. The study will test whether the direct effects of LTPA are relatively larger in older subjects (>=75 years) compared to younger subjects, and if the direct effects are larger in those with diabetes and underlying CVD.

Materials and Methods

Study Sample

Study participants were 947 women and 708 men (n=1655) who were a subset of 2092 participants 55 years and older who resided in and around Sonoma, California and who participated in a community-based, longitudinal study on aging. The protocol for the original study was approved by Committees for the Protection of Human Subjects, University of California, Berkeley and University of California, San Francisco. The current study of the short-term effects of LTPA on walking speed in the elderly was approved by the Committee for Protection of Human Subjects, University of California, Berkeley.

Measurements of the 1655 participants in the current study included direct assessments of bioelectric impedance, walking speed, and self-reported measures of leisure-time physical activity (LTPA) and physician-diagnosed health conditions (e.g., myocardial infarction, diabetes) at a baseline evaluation (May 1993-December 1994). Participants were re-evaluated for these various measures at each of 3 subsequent time points, separated by approximately 18 month intervals. In addition to these measures, subjects were assessed in terms of a variety of characteristics (See Study Measures section below). Data were available for 1283, 952, and 674 participants at the 3 follow-up periods, respectively. Subjects who did not complete the study were lost to follow-up due to death (N=102), relocation (N=3), refusal/inability to complete bioimpedance measurement protocol (N=354), or refusal/inability to complete other portions of the study protocol (N=336).

Study Measures

Measurement of body composition

Estimates of fat mass and lean mass were derived from measurements of bioelectric impedance (BI) based equations developed from simultaneous measurements of dual-energy x-ray absorptiometry (DEXA) and BI in a subsample of 200 study subjects. Details are given in a previous chapter (See Chapter 3). A relative measure of

lean mass to fat mass (lean:fat mass ratio) was computed for purposes of the current study.

Measurement of Leisure-Time Physical Activitiy (LTPA)

Subjects were assessed with respect to their average weekly participation in 22 activities in the year prior to interview. Energy expenditure was derived based on frequency of participation and the metabolic equivalence (METS) of these different activities based on a standard compendium (32). Categories of energy expenditure were created based on public health recommended guidelines (0 Mets/week, 0-22.5 Mets/week, >22.5 – 35 Mets/week, >35 Mets/week) (33). Additional details of the LTPA measure are provided in Chapter 3.

Diabetes and other cardiovascular disease

Self-reported diabetes and cardiovascular disease (CVD) were physician-diagnosed conditions and reported vascular surgery or angioplasty on coronary, peripheral, and cerebral arteries. Cerebrovascular disease included stroke and TIA, in addition to surgery. Cardiac disease included MI, angina pectoris, congestive heart failure, and serious arrhythmias. Year of first diagnosis was reported at the baseline evaluation. Incident cases –i.e., new diagnoses of diabetes and CVD since previous interview-were reported at all follow-up evaluation periods.

Walking Speed

Walking speed was measured by number of times study participants walked around a 10 foot length of rope for 60 seconds (i.e., number of lengths*10ft/60 sec). Study participants were allowed to use a walking aids (e.g., cane or walker), if needed, to complete the task.

Other covariates

Measures of body mass index (>25-30 kg/m^2, >30 kg/m^2) was defined as weight divided by $height^2$. Waist circumference was included as a measure of fat distribution. At each survey, subjects were classified as having CVD as well as other health-related problems (e.g., asthma, COPD, cancer, kidney or liver disease, Parkinson's disease). Depression was based on Center for Epidemiologic Studies Depression (CES-D) scale $\geq$ 16, and/or use of antidepressant medications. Smoking history was defined as never (i.e., no reported history of cigarette smoking, or <10 pack-years and having quit more than 20 years prior to the study), current, and ex-smoker(i.e., no current reported smoking and $\geq$10 pack-years or having quit $\leq$ 20 years ago prior to study). Education was assessed as a 3-level variable: < equivalent of a high school (H.S.) diploma, H.S. diploma, or > H.S.

diploma. Income was categorized as having an annual income of <20K, 20-39K, 40-75K, ≥ 75K. Subjects rated their overall health as excellent/good or fair/poor. Living arrangements were defined as living alone, living with a spouse, or living with a non-spouse. Physical function was assessed as reported difficulty to carry out one or more physical tasks (e.g., carrying a bag of groceries over 10lbs). Additional details with regard to physical function are provided in Chapter 3.

Analysis

Descriptive Analysis

An initial analysis was undertaken to characterize walking speed in men and women separately. Level differences and changes in walking speed over the 8-year study were examined in preliminary models to assess the variability of the measure. Plots of various body composition measures—e.g., lean mass, fat mass, lean:fat (Lnfat) mass ratio, and waist circumference-- were examined with walking speed (both unadjusted and age-adjusted). Other plots included walking speed and the different categories of LTPA, as well as diabetes, BMI, depression, CVD. Various covariates were selected for the analysis given their association with walking speed (see section 'Other Covariates' above), and accounted for as potential confounders in the causal inference analysis described below.

Given its distribution in relation to walking speed, Lnfat was dichotomized as high vs. low, based on its median value in women (≤1.50, >1.50) and men (≤2.65, 2.65), respectively (See plots in Appendix 4). LTPA was dichotomized as high vs. low based on ≥22.5 Mets/week in both men and women, which is the minimal weekly recommended level of physical activity for the adult population (i.e., energy expenditure equivalent to 30 minutes of brisk walking/5 days per week) (33). Further details of the LTPA measure are provided in Chapter 3.

Causal Inference Analysis

An analysis based on history-adjusted marginal structural models (HAMSMs) was undertaken to investigate the potential direct effects of LTPA on walking speed independently of Lnfat (see Figure 1 in previous section). Models were developed for walking speed for each evaluation period *t* in the study *(t=1,…,4)*, to assess the short-term effects of LTPA with respect to walking speed for fixed levels of Lnfat in the population. Targeted Maximum Likelihood estimation (TMLE) was used to estimate the effects of LTPA (based on reported activity in the year prior to interview at *t*) in the presence of Lnfat (analysed as a causal intermediate on the pathway between LTPA and walking speed) and other covariates (analysed as confounders of the LTPA and Lnfat associations with walking speed) at each *t*. (Details are given in the Appendix).. The results of the different analyses for each *t (t=1,…,4)* were examined and pooled. Mean estimates of walking speed were obtained for different combinations of LTPA(=0,1 < 22.5, ≥ 22.5 Mets/week, respectively) and Lnfat(=0,1, ≤ median, > median, respectively) which represented targeted causal parameters in the modeling procedure. Based on these estimates, direct effects of LTPA on walking speed were computed from the difference in

means at fixed levels of Lnfat---i.e., the direct effects of LTPA if everyone in the population had greater or lesser Lnfat.

Given the implications of different walking speeds for differences in distance walked over time, estimates of mean walking speed were converted to distance walked based on an assumed constant rate of walking speed multiplied by different fixed times (i.e., 15min, 30min, 60min). Mean distance walked for different combinations of LTPA(=0,1) and Lnfat(=0,1) were compared to examine direct effects of LTPA on ability to walk greater distances.

Stratified analyses were carried out to determine if the direct effects of LTPA on walking speed, that accounted similarly for Lnfat on the causal pathway, differed according to age, diabetes or CVD status. The analysis was conducted for each time point *t*, at which these stratification variables were updated (e.g., incident diabetes) in order to update the effects of LTPA on walking speed that could differ by subjects' age (< 75, ≥ 75 years), diabetes status (yes/no), and presence of cardiovascular disease (yes/no) at the beginning of each interval.

Analyses were carried out with SAS version 9.1.3 and R software version 2.4.1.

Results

Characteristics of study factors and walking speed from the study's baseline assessment are provided in Table 1. Mean walking speed was comparable in men and women (2.29 vs. 2.27 ft/sec), with 80% of the distribution (10^{th}, 90^{th} percentiles) in both sexes between 1.67 and 2.83 ft/sec. Men participated in greater levels of physical activity (77.2% reported $\geq$22.5 Mets/week compared to 70.7% of women) but experienced more health-related problems—e.g., a higher percentage of men than women (7.2% vs. 3.8%) reported diabetes and CVD (23.6% vs. 11.1%) at the baseline interview. Conversely, women had on average a lower lean:fat mass ratio than men (1.50 vs 2.65), had lower mean BMI ($\geq$ 25-30 kgm^{-2}: 36.6% vs. 52.6%; $\geq$30 kgm^{-2}: 18.8% vs. 21.0%), and more physical functioning difficulties (28.9% vs. 15.3%).

Table 1A. Baseline Characteristics and Walking Speed in Women

	N (%)*	Mean‡	SD		N (%)	Mean	SD
LTPA				C-ESD>16			
< 22.5	273 (29.3)	2.11	0.51	Yes	98	2.16	0.52
≥ 22.5	659 (70.7)	2.34	0.50	No	798	2.30	0.50
				Missing	34	2.03	0.53
Lnfat							
≤ 1.50	509 (54.6)	2.24	0.48	Antidepressant			
> 1.50	424 (45.4)	2.32	0.54	Yes	52 (5.6)	2.09	0.57
				No	880 (94.4)	2.28	0.51
Age							
< 75	698 (74.9)	2.39	0.47	Smoking History			
≥ 75	234 (25.1)	1.93	0.47	Never	455 (48.8)	2.30	0.53
				Current	81 (8.7)	2.24	0.50
Diabetes				Former	396 (42.5)	2.26	0.49
Yes	35 (3.8)	2.29	0.51				
No	897 (96.2)	1.89	0.50	Education			
				<HS	79 (8.5)	2.00	0.53
Health†				HS	259 (27.8)	2.25	0.47
None	532 (57.5)	2.38	0.47	>HS	594 (63.6)	2.32	0.52
CVD	103 (11.1)	2.01	0.54				
Other	288 (31.4)	2.17	0.53	Income			
				<20K	210	2.07	0.53
BMI				≥ 20-40	337	2.27	0.45
0-25	416 (44.6)	2.31	0.55	≥ 40-75	235	2.40	0.53
>25-30	341 (36.6)	2.29	0.46	≥75K	67	2.50	0.46
>30	175 (18.8)	2.15	0.49	Missing	82	2.24	0.51
Physical Limitation				Living Arrangement			
Yes	269 (28.9)	1.99	0.51	Alone	370 (39.7)	2.17	0.52
No	663 (71.1)	2.38	0.47	Spouse	497 (53.3)	2.35	0.49
				Other	65 (7.0)	2.23	0.54

*N (%) represent frequency and proportion of total N in each category. Mean,SD represent mean and standard deviation of walking speed in each category.

†Health was missing for 9 subjects

‡Range of 0-7 missing values for speed (e.g., unable/refused) for different categories.

Table 1B. Baseline Characteristics and Walking Speed in Men

	N (%)*	Mean‡	SD		N (%)	Mean	SD
LTPA				C-ESD>16			
< 22.5	158 (22.8)	2.14	0.49	Yes	30 (4.3)	2.10	0.57
≥ 22.5	536 (77.2)	2.33	0.46	No	639 (92.1)	2.31	0.46
				Missing	25 (3.6)	2.01	0.56
Lnfat							
≤ 2.65	348 (51.1)	2.24	0.45	Antidepressants			
> 2.65	346 (49.9)	2.33	0.49	Yes	18 (2.6)	2.04	0.46
				No	676 (97.4)	2.29	0.47
Age							
<75	520 (73.5)	2.38	0.44	Smoking History			
>75	184 (26.5)	2.02	0.46	Never	232 (33.4)	2.35	0.47
				Current	46 (6.6)	2.30	0.47
Diabetes				Former	416 (59.9)	2.25	0.47
Yes	50 (7.2)	2.18	0.47				
No	644 (92.8)	2.29	0.47	Education			
				<HS	73 (10.5)	2.10	0.49
Health†				HS	107 (15.4)	2.19	0.40
None	371 (53.7)	2.38	0.44	>HS	514 (74.1)	2.33	0.48
CVD	162 (23.6)	2.12	0.47				
Other	156 (22.7)	2.23	0.49	Income			
				<20K	54 (7.8)	2.08	0.46
BMI				20-40	234 (33.7)	2.19	0.46
0-25	183 (26.4)	2.30	0.49	40-75	248 (35.7)	2.33	0.46
>25-30	365 (52.6)	2.31	0.47	>75K	119 (17.2)	2.50	0.42
>30	146 (21.0)	2.21	0.43	Missing	39 (5.6)	2.18	0.52
Physical Limitation				Living Arrangement			
Yes	106 (15.3)	1.94	0.44	Alone	81 (11.7)	2.18	0.51
No	588 (84.7)	2.35	0.45	Spouse	580 (83.5)	2.30	0.46
				Other	33 (4.8)	2.27	0.51

*N (%) represent frequency and proportion of total N in each category. Mean,SD represent mean and standard deviation of walking speed in each category.
†Health was missing for 5 subjects
‡Range of 0-9 missing values for speed (e.g., unable/refused) for different categories.

Based on models of the effects of LTPA and Lnfat at each evaluation period, population-level mean estimates of walking speed were obtained (See Table 2). These estimates reflect the population-level mean walking speed if, contrary to fact, subjects participated in levels of LTPA less than vs. greater or equal to 22.5 Mets/week, and, if contrary to fact, had relatively greater (> sex-specific median Lnfat) or lesser (≤ sex-specific median Lnfat) lean mass to fat mass. In addition to overall, population-level mean estimates of walking speed, subpopulation-level mean estimates were determined, based on stratified levels of age, diabetes, and CVD health status obtained for each time point.

Table 2 indicates an increase in mean walking speed if contrary to fact everyone experienced higher Lnfat and LTPA at each time. In women at time 1, the overall mean difference in walking speed if everyone participated at higher LTPA and Lnfat was 2.354 ft/sec vs. 2.157 ft/sec, if contrary to fact, everyone had lower or no LTPA and lower Lnfat. Overall population-level results in the men indicated a similar pattern.

In the stratified analysis, mean walking speed was lower for the subpopulation of subjects $\geq$ 75 years; those with reported diabetes; and those with CVD. A similar pattern of increased mean walking speed was observed in these subpopulations, for increased LTPA and Lnfat, as seen in the overall analysis. In fact, there was no indication of stratum-specific mean differences in walking speed for different LTPA and Lnfat that was observed for either the men or the women. For example, in men <75 years at time 1, the mean walking speed, if everyone participated at higher levels of LTPA but had lower Lnfat, was 2.354 ft/sec vs. 2.252 ft/sec if everyone had lower LTPA and lower Lnfat(i.e. mean difference 2.354 ft/sec-2.252 ft/sec = 0.102 ft/sec). In men >75, although the means were lower (2.001 ft/sec vs. 1.898 ft/sec), the mean differences based on the same counterfactual comparison of LTPA and Lnfat was approximately the same (i.e. mean difference 2.001 ft/sec-1.898 ft/sec = 0.103 ft/sec). Based on these stratified analyses, there was no evidence to indicate that the estimated direct effects of LTPA on walking speed might vary in the subpopulations that were examined.

Table 2A. Stratified Analysis-Mean Walking Speed (Ft/Sec) in Females based on Targeted Maximum Likelihood Estimation.

Overall

	Met0,Lnfat0*	Met0,Lnfat1	Met1,Lnfat0	Met1,Lnfat1
Time 1	2.157	2.243	2.268	2.354
Time2	2.261	2.350	2.305	2.394
Time 3	2.184	2.239	2.311	2.366
Time 4	2.417	2.460	2.512	2.555

Age Stratified

	Age <75				Age >=75			
	Met0,Lnfat0	Met0,Lnfat1	Met1,Lnfat0	Met1,Lnfat1	Met0,Lnfat0	Met0,Lnfat1	Met1,Lnfat0	Met1,Lnfat1
Time 1	2.265	2.352	2.376	2.462	1.828	1.915	1.940	2.026
Time2	2.407	2.496	2.450	2.539	1.924	2.012	1.967	2.056
Time 3	2.332	2.386	2.459	2.514	1.938	1.992	2.065	2.119
Time 4	2.602	2.645	2.696	2.740	2.176	2.219	2.270	2.313

Diabetes Stratified

	No				Yes			
	Met0,Lnfat0	Met0,Lnfat1	Met1,Lnfat0	Met1,Lnfat1	Met0,Lnfat0	Met0,Lnfat1	Met1,Lnfat0	Met1,Lnfat1
Time 1	2.171	2.258	2.282	2.369	1.790	1.874	1.902	1.985
Time2	2.273	2.362	2.317	2.405	2.035	2.124	2.078	2.167
Time 3	2.200	2.254	2.327	2.381	1.948	2.003	2.076	2.130
Time 4	2.429	2.472	2.524	2.567	2.224	2.267	2.319	2.362

CVD Stratified

	No				Yes			
	Met0,Lnfat0	Met0,Lnfat1	Met1,Lnfat0	Met1,Lnfat1	Met0,Lnfat0	Met0,Lnfat1	Met1,Lnfat0	Met1,Lnfat1
Time 1	2.174	2.261	2.285	2.372	2.014	2.100	2.125	2.211
Time2	2.287	2.376	2.330	2.419	2.079	2.168	2.122	2.211
Time 3	2.225	2.279	2.352	2.407	2.001	2.056	2.128	2.183
Time 4	2.463	2.506	2.558	2.601	2.177	2.221	2.272	2.315

*Met0 <22.5Mets/wk Met1 ≥22.5 Mets/wk Lnfat0 ≤median Lnfat1 >median

Table 2B. Stratified Analysis-Mean Walking Speed (Ft/Sec) in Males based on Targeted Maximum Likelihood Estimation.

Overall

	Met0,Lnfat0*	Met0,Lnfat1	Met1,Lnfat0	Met1,Lnfat1
Time 1	2.160	2.267	2.262	2.370
Time2	2.310	2.382	2.393	2.465
Time 3	2.251	2.278	2.305	2.332
Time 4	2.456	2.485	2.536	2.564

Age Stratified

	Age <75				Age >=75			
	Met0,Lnfat0	Met0,Lnfat1	Met1,Lnfat0	Met1,Lnfat1	Met0,Lnfat0	Met0,Lnfat1	Met1,Lnfat0	Met1,Lnfat1
Time 1	2.252	2.359	2.354	2.462	1.898	2.005	2.001	2.108
Time2	2.445	2.517	2.528	2.600	2.030	2.102	2.113	2.185
Time 3	2.369	2.396	2.423	2.450	2.092	2.119	2.146	2.173
Time 4	2.617	2.645	2.697	2.725	2.278	2.306	2.357	2.385

Diabetes Stratified

	No				Yes			
	Met0,Lnfat0	Met0,Lnfat1	Met1,Lnfat0	Met1,Lnfat1	Met0,Lnfat0	Met0,Lnfat1	Met1,Lnfat0	Met1,Lnfat1
Time 1	2.163	2.270	2.265	2.373	2.121	2.229	2.224	2.331
Time2	2.319	2.391	2.401	2.473	2.212	2.284	2.295	2.367
Time 3	2.264	2.292	2.318	2.345	2.114	2.142	2.168	2.196
Time 4	2.476	2.505	2.556	2.584	2.286	2.314	2.366	2.394

CVD Stratified

	No				Yes			
	Met0,Lnfat0	Met0,Lnfat1	Met1,Lnfat0	Met1,Lnfat1	Met0,Lnfat0	Met0,Lnfat1	Met1,Lnfat0	Met1,Lnfat1
Time 1	2.193	2.300	2.296	2.403	2.054	2.161	2.157	2.264
Time2	2.369	2.441	2.452	2.524	2.129	2.201	2.211	2.283
Time 3	2.303	2.331	2.357	2.384	2.121	2.148	2.174	2.202
Time 4	2.513	2.540	2.592	2.620	2.312	2.340	2.392	2.420

*Met0 <22.5Mets/wk Met1 ≥22.5 Mets/wk Lnfat0 ≤median Lnfat1 >median

Given that the mean differences in walking speed based on the stratified analysis were essentially the same as for the overall analysis, for both the men and women, further analysis was conducted with the overall population-level estimates of walking speed only. Estimates were pooled over the 4 evaluation periods and compared based on the different counterfactuals (e.g., high LTPA, high Lnfat vs. low LTPA, low Lnfat) that could hypothetically have occurred in this population (Table 3).

Table 3 Pooled Analysis-Overall Mean Estimates of Walking Speed in Women and Men based on Targeted Maximum Likelihood Estimation.

	Counterfactual*	Walking Speed Mean (ft/sec)	95% Confidence Limits	
			Lower 95	Upper 95
Females	Met0, Lnfat0	2.238	2.113	2.363
	Met0, Lnfat1	2.316	2.191	2.441
	Met1,Lnfat0	2.316	2.191	2.441
	Met1,Lnfat1	2.394	2.269	2.519
Males	Met0, Lnfat0	2.278	2.126	2.430
	Met0, Lnfat1	2.345	2.193	2.497
	Met1,Lnfat0	2.326	2.174	2.478
	Met1,Lnfat1	2.418	2.266	2.570

*Met0 <22.5Mets/wk Met1 ≥22.5 Mets/wk Lnfat0 ≤median Lnfat1 >median

In women, based on pooled the data, mean walking speed increased with higher LTPA and greater LNFAT. Mean walking speed was lowest, if contrary to fact, everyone had LTPA<22.5 Mets/week and < median Lnfat (2.238 ft/sec; 95% CI 2.113, 2.363); and, conversely, mean walking speed was highest, if contrary to fact, everyone had LTPA>22.5 Mets/week and > median Lnfat (2.394 ft/sec; 95% CI 2.269, 2.519). These mean values were significantly different ($p<0.05$). Effect of higher LTPA alone (i.e., direct effect of LTPA) on walking speed indicated a non-significant increase in mean walking speed if everyone was of low Lnfat (2.316 vs 2.238 ft/sec), or high Lnfat (2.394 vs. 2.316 ft/sec). Similar non-significant increases in walking from direct effects of LTPA alone were observed for the men.

Results that describe the differences in total distance traversed given the different mean walking speed estimates from the previous table are provided in Table 4.

Table 4. Pooled Analysis—Overall Mean Estimates of Distance Walked* in Women and Men based on Targeted Maximum Likelihood Estimation.

	Counterfactual†	Mean Difference‡	95% Confidence Limits	
			Lower 95	Upper 95
Females	Met1,Lnfat0-Met0,Lnfat0			
Time walked	15min	70	-42	183
	30min	140	-85	365
	60min	281	-170	731
	Met1,Lnfat1-Met0,Lnfat1			
	15min	70	-42	183
	30min	140	-85	365
	60min	280	-170	731
	Met1,Lnfat1-Met0,Lnfat0			
	15min	140	28	253
	30min	281	56	506
	60min	561	111	1012
			Lower 95	Upper 95
Males	Met1,Lnfat0-Met0,Lnfat0			
Time walked	15min	43	-94	180
	30min	86	-188	360
	60min	172	-375	719
	Met1,Lnfat1-Met0,Lnfat1			
	15min	66	-71	203
	30min	131	-142	405
	60min	263	-284	810
	Met1,Lnfat1-Met0,Lnfat0			
	15min	126	-11	263
	30min	252	-22	525
	60min	503	-44	1051

Mean distance walked (in feet) for time given (in minutes) is defined as mean walking speed (ft/sec)(min)*(60sec/min)

†Met1 $\geq$22.5 Met/wk Met0 <22.5Met/wk Lnfat1>median Lnfat0$\leq$median

‡Mean difference can be interpreted as the mean distance that separates a population based on one set of counterfactuals compared with another.

These results provide a sense of the differences in walking distance the population could have experienced based on different hypothetical counterfactual scenarios (high, low LPTA, Lnfat). The distance traversed by subjects was greater if, contrary to fact, subjects had LTPA $\geq$ 22.5Mets/week and > median Lnfat than if the same subjects did not. In women, for example, the calculated differences for high LTPA and Lnfat vs. low LTPA and Lnfat were: 140ft, 281ft, and 561feet after 15, 30 and 60 minutes of walking, respectively. For the men, based on the same comparison of LTPA and Lnfat as the women, the differences in distance walked were somewhat less (i.e., 126, 252, and 503 feet, respectively) and more imprecisely estimated than for the women.

Walking distances based on an examination of the counterfactuals to evaluate potential direct effects of LTPA differed somewhat between men and women. In the women, the distances traversed if everyone had $\geq$ 22.5 Mets/week LTPA vs. < 22.5 Mets/week were the same for different high and low Lnfat (i.e., 70, 140, and 280ft, after 15, 30, and 60 minutes of walking, respectively). However, if the same comparison of LTPA was made, the men whose Lnfat was higher performed somewhat better (i.e., 66, 131, 263ft) than if their Lnfat was lower (i.e., 43, 86, and 172ft). This latter result may reflect an underlying physiologic difference in the effects of LTPA for different levels of Lnfat in males (e.g., improved aerobic capacity with LTPA with leanness), or inherent differences in the LTPA measure (i.e., amount and/or type) for the different levels of Lnfat.

Although no significant difference mean distance walked was observed based on direct effects of LTPA on walking speed in Table 4, the distances were greater nonetheless and suggested a potential direct benefit of LTPA for preservation of walking speed.

Discussion

Estimates of the short-term effects of LTPA, reported in the prior year, on walking speed suggested that > 22.5 Mets/week of LTPA could contribute to higher mean walking speed, in sex-specific analyses; however, the results were not measured with enough precision to draw any conclusions that this was indeed the case. Moreover, estimates based on a stratified analysis did not indicate that short-term direct effects of LTPA differed by age, diabetes, or CVD status, as hypothesized. However, the effect of $\geq$ 22.5 Mets/week of LTPA and > median Lnfat combined to produce greater mean walking speed with greater precision, for both women and men, than if contrary to fact, individuals from either of these groups had < 22.5 Mets/week and $\leq$ median Lnfat.

The fact that direct effects of LTPA as estimated contributed only marginally to higher walking speed should not be overlooked: small differences in walking speed could ultimately mean larger differences with respect to distance covered over time, as results of this study indicated, though not with precision. The amount of distance traversed by an individual in a given amount of time could represent a measure of that person's lower body function. In this regard, LTPA could have broader implications beyond its influence on energy utilization and metabolism with regard to preservation of lower-body function.

The fact that the stratified analysis did not show any increased direct effects of LTPA in groups where it was expected the influence of LTPA on walking speed would be relatively greater compared to its indirect effects through Lnfat, could be due to

several factors. The stratified groups may have not been sufficiently different from one another. It is possible that the subjects' who remained in the study who were older, or who received a diagnosis of diabetes or CVD, were relatively 'healthy' compared to those who did not, since the former subjects represent a survivor population. Also, a greater proportion of the subjects who were counted in the analysis as >75 years, or with a diagnosis of diabetes or CVD, may have become so during follow-up, which again could blur any distinction with individuals in the 'healthy' strata (i.e., <75 year, no diabetes, no CVD). Still, their mean walking speed estimates were considerably lower than for the 'healthy' group. Censoring weights were used in the analysis to account for the selection bias that could have contributed to these results. However, estimates based on analyses with and without the weights did not change substantially. The censoring models from which these weights were derived included variables like age and health, but may have not been sensitive enough to account for other survivor group attributes.

It is also possible that direct effects of LTPA did not differ in the stratified analysis because the LTPA $\geq$ 22.5 Mets/ week reported by individuals may have been of a sufficient level so that most of the observed effects of walking speed were through Lnfat rather than other pathways. Frontera *et al.* examined the effects of an exercise program on muscle mass and muscle strength in men older than 75 years, and found that with high levels of physical activity, muscle mass and muscle strength increased (34). Participants in the current study who reported $\geq$ 22.5 Mets/week were comprised mostly of individuals who reported > 35 Met/week, particularly among the men. Therefore, it is possible that the levels of LTPA reported by this study's subjects on the whole was of sufficient magnitude to reverse aging processes associated with less efficient energy utilization (i.e., improved activity of oxidative enzymes, improved glucose transportation, improved mitochondrial function) (19, 30, 31). If so, the observed direct effects of LTPA would have been relatively smaller than its indirect effects through Lnfat. This may in part be the reason, too, for the small, imprecisely measured direct effects of LTPA in the overall analysis, particularly in the men.

In terms of some of the study measures, dichotomized forms of LTPA and Lnfat were used for this study. This was done partly because of the distributions of these variables relative to walking speed, but dichotomized forms were also chosen in order to ensure there were enough individuals in each of the different groups to perform the stratified analysis. For example, if Lnfat had been categorized more finely, it would have not been possible to examine effects by diabetes status, given there were no individuals with diabetes and higher Lnfat. A more objective measure of diabetes that did not depend on self-report, but instead was based on fasting glucose levels, may have been a better surrogate measure to test whether the effects of LTPA differed by diabetes status. By contrast, Lnfat, which was derived as an objective measure of body composition, provided a useful surrogate of metabolic intermediary processes by which the direct effects of LTPA on walking speed could be assessed.

In summary, the study found that walking speed improved with greater LTPA and Lnfat. In the overall analysis (i.e., overall populations) of men and women, the results suggested there was some improvement in walking speed as a result of the direct influence of LTPA (i.e., not explained by body composition); however, these effects were imprecisely measured. Evidence based on the overall analysis, and results from the stratified analysis, suggest that the relative effects of LTPA on walking speed are greater

through Lnfat than through other mechanisms, which the 'direct' effects were intended to represent. These findings underscore the importance of LTPA on metabolism and energy utilization that affect mobility and lower-body function in the elderly.

References

1. Shinkai S, Watanabe S, Kumagai S, et al. Walking speed as a good predictor for the onset of functional dependence in a Japanese rural community population. Age Ageing 2000;29:441-6.
2. Newman AB, Simonsick EM, Naydeck BL, et al. Association of Long-Distance Corridor Walk Performance With Mortality, Cardiovascular Disease, Mobility Limitation, and Disability. JAMA 2006;295:2018-2026.
3. Brach JS, Studenski SA, Perera S, VanSwearingen JM, Newman AB. Gait variability and the risk of incident mobility disability in community-dwelling older adults. J Gerontol A Biol Sci Med Sci 2007;62:983-8.
4. Weiss CO, Fried LP, Bandeen-Roche K. Exploring the hierarchy of mobility performance in high-functioning older women. J Gerontol A Biol Sci Med Sci 2007;62:167-73.
5. Centers for Disease Control. Physical Activity and Health: A Report of the Surgeon General. Pittsburgh: Dept. of Health and Human Services--S/N017-123-00196-5, 1996.
6. Haskell WL, Lee IM, Pate RR, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Circulation 2007;116:1081-93.
7. Nelson ME, Rejeski WJ, Blair SN, et al. Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007;39:1435-45.
8. Paffenbarger RS, Jr,, Hyde RT, Wing AL, Hsieh C-C. Physical activity, all-cause mortality, and longevity of college alumni. New England Journal of Medicine 1986;314:605-13.
9. Brooks G. Exercise Physiology: McGraw-Hill, 2005.
10. Lee IM. Dose-response relation between physical activity and fitness: even a little is good; more is better. Jama 2007;297:2137-9.
11. Lee I. Epidemiologic Methods in Physical Activity Studies, 2009.
12. Mazzeo R, Cavanagh P, Evans WJ, et al. Exercise and Physical Activity for Older Adults. Medicine and Science in Sports and Exercise 1998;30.
13. Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. J Gerontol A Biol Sci Med Sci 2000;55:M221-31.
14. Dumurgier J, Elbaz A, Ducimetiere P, Tavernier B, Alperovitch A, Tzourio C. Slow walking speed and cardiovascular death in well functioning older adults: prospective cohort study. Bmj 2009;339:b4460.

15. Frontera WR, Suh D, Krivickas LS, Hughes VA, Goldstein R, Roubenoff R. Skeletal muscle fiber quality in older men and women. American Journal of Physiology and Cell Physiology 2000;279:C611-8.
16. Vandervoort A. Aging of the Human Neuromuscular System. Muscle Nerve 2002;25:17-25.
17. Elbaz A, Ripert M, Tavernier B, et al. Common carotid artery intima-media thickness, carotid plaques, and walking speed. Stroke 2005;36:2198-202.
18. Rosano C, Brach J, Longstreth Jr WT, Newman AB. Quantitative measures of gait characteristics indicate prevalence of underlying subclinical structural brain abnormalities in high-functioning older adults. Neuroepidemiology 2006;26:52-60.
19. Williams GN, Higgins MJ, Lewek MD. Aging skeletal muscle: physiologic changes and the effects of training. Phys Ther 2002;82:62-8.
20. Sowers MR, Crutchfield M, Richards K, et al. Sarcopenia Is Related to Physical Functioning and Leg Strength in Middle-Aged Women. J Gerontol A Biol Sci Med Sci 2005;60:486-490.
21. Reid KF, Naumova EN, Carabello RJ, Phillips EM, Fielding RA. Lower extremity muscle mass predicts functional performance in mobility-limited elders. J Nutr Health Aging 2008;12:493-8.
22. Sternfeld B, Ngo L, Satariano WA, Tager IB. Associations of body composition with physical performance and self-reported functional limitation in elderly men and women. American Journal of Epidemiology 2002;156:110-21.
23. Estrada M, Kleppinger A, Judge JO, Walsh SJ, Kuchel GA. Functional impact of relative versus absolute sarcopenia in healthy older women. J Am Geriatr Soc 2007;55:1712-9.
24. Gregg EW, Brown A. Cognitive and Physical Disabilities and Aging-Related Complications of Diabetes. Clinical Diabetes 2003;21:113-118.
25. Aalami OO, Fang TD, Song HM, Nacamuli RP. Physiological features of aging persons. Arch Surg 2003;138:1068-76.
26. Muller DC, Elahi D, Tobin JD, Andres R. The effect of age on insulin resistance and secretion: a review. Semin Nephrol 1996;16:289-98.
27. Lewis GF, Carpentier A, Adeli K, Giacca A. Disordered fat storage and mobilization in the pathogenesis of insulin resistance and type 2 diabetes. Endocr Rev 2002;23:201-29.
28. Goodyear LJ, Kahn BB. Exercise, glucose transport, and insulin sensitivity. Annu Rev Med 1998;49:235-61.
29. Petersen KF, Befroy D, Dufour S, et al. Mitochondrial dysfunction in the elderly: possible role in insulin resistance. Science 2003;300:1140-1142.
30. Kim HJ, Lee JS, Kim CK. Effect of exercise training on muscle glucose transporter 4 protein and intramuscular lipid content in elderly men with impaired glucose tolerance. Eur J Appl Physiol 2004;93:353-8.
31. Lanza IR, Short DK, Short KR, et al. Endurance exercise as a countermeasure for aging. Diabetes 2008;57:2933-42.
32. Ainsworth BE, Haskell WL, Leon AS, et al. Compendium of physical activities: Classification of energy costs of human physical activities. Medicine and Science in Sports and Exercise 1993;25:71-80.

33. Pate RR, Pratt M, Blair SN, et al. Physical activity and public health: A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. Journal of the American Medical Association 1995;273:402-7.
34. Frontera WR, Meredith CN, O'Reilly KP, Knuttgen HG, Evans WJ. Strength conditioning in older men: skeletal muscle hypertrophy and improved function. Journal of Applied Physiology 1988;64:1038-44.

Conclusion

Recent development of causal inference methods to quantify causal effects based on observational data has led to their growing application in epidemiologic research. These tools provide estimates of population-level (i.e., marginal) effects which are of more direct interest than conditional effects based on conventional association models, which have predominated in epidemiologic data analyses. Moreover, these tools allow for estimation of effects in the presence of time-dependent confounding, which represents an important limitation of conventional approaches. In general, these new methods allow for the evaluation of effects in the presence of factors on causal pathways that would be difficult if not impossible to address with standard statistical approaches, and a more valid approach for the separation of direct and indirect effects of an exposure when one pathway is hypothesized to include a measurable intermediate between exposure and outcome.

The advantages of these new methods compared to standard approaches in observational studies makes them ideally suited to examine causal questions for longitudinal studies of aging. The methods are well suited to disentangle the complex interrelated temporal relationships between factors (e.g., physical function, physical activity) and allow assessment of the individual contributions of these factors on the aging process. The methods can be used to examine the contribution of factors at different points in the aging process (e.g., early-life events) given that these methods can control for health intermediates, which commonly presents an analytical dilemma with the use of conventional statistical methods. Furthermore, these causal approaches can be used to evaluate effect-modification by time-varying covariates, which is particularly relevant for the understanding the potential benefits/limits of particular effects (i.e., interventions), given changes in underlying health characteristics in aging populations over time.

Causal statistical models (e.g., MSMs, HAMSMs) were applied, and the properties inherent to causal methods were evoked, to address three questions of aging in the analyses included in this dissertation.

In the first study, effects of leisure-time physical activity (LTPA) and body composition (i.e., lean:fat mass ratio-L/F) were assessed to examine their individual contributions to long-term patterns of physical function. Higher L/F was found to contribute to an overall lower probability of long-term physical limitation. Higher levels of LTPA were found to reduce the risk of new physical limitations (i.e., recurrent limitation in recovered subject or incident limitation), conditional on the level functioning (e.g., high vs. low) conferred by L/F. In future research, other comparisons could be made between different factors to elucidate roles of these on important aging outcomes.

In another analysis, LTPA was examined with respect to different pathways by which it could contribute to walking speed. The analysis accounted for the likely indirect pathway (i.e., effect) of LTPA through L/F, and summed its contribution through all other pathways as its 'direct' effect on walking speed. This direct effect was found to be marginally protective but was imprecisely measured. Moreover, the direct effect of LTPA relative to its indirect effect through L/F did not differ across groups with suspected or evident metabolic derangements (e.g., diabetes), for which the direct effect of LTPA was

expected to be larger. This work suggested that the effects of LTPA on metabolism (e.g., glucose utilization, cellular oxidative processes), that ultimately affect mobility and function, continue throughout the aging process.

In the last analysis, lifetime secondhand smoke exposure (SHS) was examined with respect to its potential independent health effects on dementia risk. The analysis accounted for the presence of clinical CVD—a factor on the causal pathway that would potentially mediate the effects of early life SHS exposure and dementia risk. The results did not find that SHS alone increased the risk of dementia, but found that there was a three-fold increase in the risk in those with high lifetime exposure (>25 years) and who had >25% carotid artery stenosis—a subclinical measure of CVD-- than subjects who had no/low SHS exposure and $\leq$25% stenosis of the carotid artery. The implications of this study are important for different reasons. The study suggests that seniors with undiagnosed CVD and with accumulated SHS exposure may be at an increased risk of developing dementia. The methodological implications of the study underscore: 1) the use of MSMs to evaluate effects of interest; and 2) the use of subclinical biomarkers of disease to provide more sensitivity than clinical measures (e.g., stroke) to examine the relationship of particular exposures (e.g., SHS) with disease.

In conclusion, the methods for causal inference that have been described in this dissertation were used to evaluate and quantify causal effects based on observational data. These methods provide an opportunity to develop more accurate causal descriptions of the different factors that are involved in aging and disablement. Improved examination and estimation of these factors, in the context of other age-related changes, are essential for understanding the role of these factors on long-term health and functioning in the elderly.

Appendices

Appendix 1 A cross-validation deletion-substitution-addition model selection algorithm: Application to marginal structural models

1. Introduction

In recent years, epidemiologists' knowledge about the theory and application of Marginal Structural Models (MSMs) to examine causal effects in observational studies has grown substantially. MSMs provide unbiased estimates of marginal effects in the presence of both causal intermediates in point treatment (exposure) studies and time-dependent confounding in longitudinal studies (1). Conventional (conditional) association models provide stratum-specific effects which are typically biased in these situations. MSMs eliminate the need to adjust for confounding in the models themselves. Instead, nuisance parameter models (e.g. treatment models) are used to address confounding, so that with MSMs one obtains a direct, unconditional assessment of the exposure on the response. While model selection procedures for nuisance parameters have been addressed in the published literature (2, 3), procedures for selection of MSMs have not. The recent development of a general cross-validated data-adaptive model selection procedure represents an important methodological advancement to better characterize the causal effects of interest through MSM selection and a more flexible examination of the exposure-response causal curve.

The cross-validation Deletion-Substitution-Addition (DSA) algorithm selects models adaptively for MSMs and nuisance parameter models for point treatment studies (4). The approach is derived from a general methodology that provides data adaptive machine learning type of algorithms based on user-supplied criteria (e.g., maximum model size) (5, 6). Specifically, the algorithm builds a model space of candidate models based on so-called deletion, substitution and addition moves and utilizes a loss-function-based estimation procedure to distinguish between different models with respect to model fit (5). The goal is to select a model that results in the best estimate of a given data distribution. Moreover, the algorithm selects models based partly on V-fold cross-validation (4, 7)and, thus, avoids the problem of "over-fitting" data that can occur with other data-adaptive model selection algorithms (e.g., StepAIC function, R-Software, current version, R Foundation for Statistical Computing).

This paper discusses methodological aspects of the algorithm and compares it with other model selection criteria. An illustrative analysis demonstrates how the algorithm works. Two R-packages are available which implement the algorithm: one of which is a well-developed package (DSA) for selection of conditional models (e.g., nuisance parameter models); the second is for MSM selection for point treatment studies (cvDSA), and includes components for the selection of nuisance parameter models (cvGLM) and selection of MSMs (cvMSM). The second package (cvDSA) is less developed than the first

in terms of ease of use and speed. We advise selection of the treatment model with the DSA package, and submission of this model to the cvMSM procedure for MSM selection. The discussion of the algorithm is in the context of its selection of MSMs, but provides an overall view of the DSA algorithm as a general tool for model selection. Both packages are available for download from http://stat-www.berkeley.edu/~laan/Software/index.html. Additional background and technical details about the algorithm are available (4-6, 8).

2. Background on MSMs

MSMs are used to define causal parameters of interest for exposure-response relations based on the concept of counterfactuals (1). This concept permits assessment of observational data in a hypothetical framework in which, contrary to fact, subjects were exposed to all possible levels of an exposure and had outcomes associated with those exposures. With counterfactual data, one can evaluate whether differences in the outcome are attributable to causal differences in the level of the exposure. To recreate the conditions under which observed data can be evaluated as counterfactual data requires several assumptions.

First, the observed data for any given subject represent one realization of his/her counterfactual data that correspond with the exposure actually received (consistency assumption) (9). In a point treatment study, the observed data can be represented as $O=(W, A, Y=Y(A))$, where W represents the baseline covariates, "A", the treatment (exposure) assignment, and $Y(A)$, the outcome under observed treatment "A". The observed data $O=(W,A,Y)$ on a randomly sampled subject represent one realization/component of the counterfactual "full" data $X=((Y(a),a\in \mathsf{A}),W)$ when exposure $a=A$.

A second assumption is the no "unmeasured confounders", or "randomization assumption": $Y(a)\perp A \mid W$ —i.e., the treatment of interest is "randomized" with respect to the outcome within strata of the measured covariates, W. (9). To satisfy this assumption, one conditions on all the measurable confounders of the exposure and outcome through a nuisance parameter model. Estimation of nuisance parameters can occur either by a model of a regression of the outcome on treatment (exposure) and all potential confounders (W) (G-computation estimation, Double Robust-Inverse Probability of Treatment Weight DR-IPTW estimation), or a model of the conditional probability of treatment given W (Inverse Probability of Treatment Weight IPTW estimation). Correct characterization of one of these nuisance parameter models is required to assess properly the effect of treatment on outcome without regard to potential extraneous factors.

Lastly, an additional assumption (experimental treatment assignment, or ETA) is required to provide unbiased estimates with IPTW estimation. This assumption states that all exposures have a positive probability of occurrence, given baseline covariates.

The parameter of interest in an MSM is the treatment-specific mean $E(Y(a)|V)$, possibly conditional on some baseline covariates V that are a subset of W ($V \subset W$). When $V=W$, the MSM represents a traditional multiple regression model, where the effect of *'a'* is a fully-adjusted causal parameter. Classical MSMs define a model for $E(Y(a)|V)$ such as a linear model $m(a,V|\beta)$, so that the parameter of interest is the regression parameter β in this assumed model. The goal of the cross-validation DSA algorithm is to achieve a correct

characterization (i.e., fit) of the nuisance models and MSM models to evaluate causal effects for point treatment studies.

Additional details of the theory and application of MSMs are available (2, 9-15).

3. Overview of the Cross-Validation Deletion, Substitution and Addition algorithm

A possible estimator of the treatment specific mean (MSM) minimizes the empirical risk -- a statistical criterion of model fit defined below -- over all candidate treatment-specific means. However, since the model space of possible treatment specific means is infinite dimensional, given the different parameterizations of the treatment variable and the baseline covariates possible in the MSM, this minimization would simply result in an over-fit model. A general solution to deal with this problem is to construct a sequence or collection of subspaces (e.g., model categories of varying size and complexity) that approximate the whole model space, a so-called sieve; then, to compute the minimizers of empirical risk for each of these subspaces. Application of V-fold cross-validation and a defined cross-validation risk criterion (described below) can be used to select the actual (optimal) subspace whose corresponding minimum empirical risk estimator minimizes the cross-validation risk (Figure 1). Given that the process described above occurs in the framework of cross-validation, with the data subset into training/test data for purposes of selecting the optimal subspace, the process is repeated, and the construction of subspaces occurs based on a whole dataset. Then the minimizer of the empirical risk in the optimal subspace selected with cross-validation becomes the final selected model.

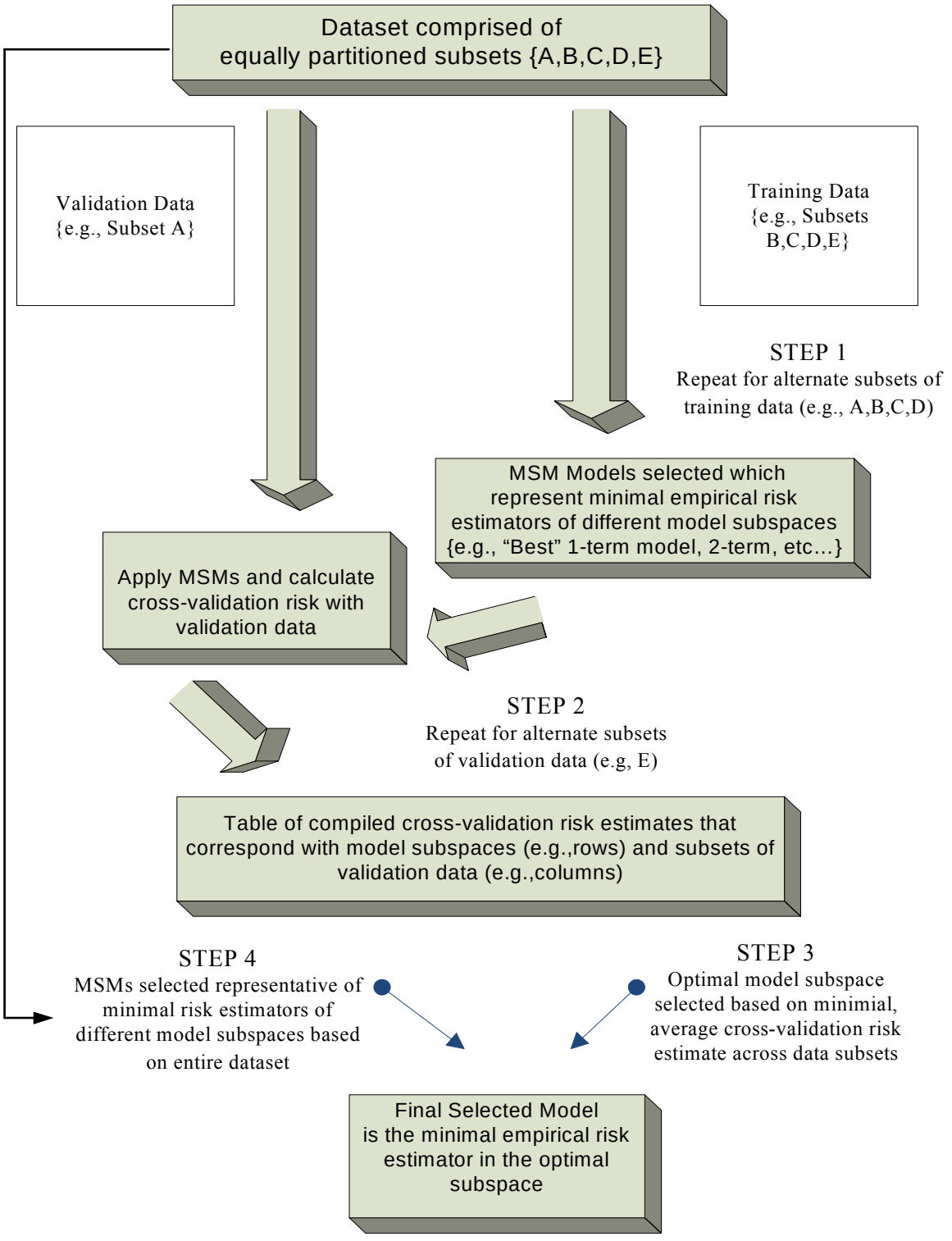


Figure 1. Overview of the Cross-validation Deletion Substitution Addition Algorithm for model selection.

4. Key methodological aspects of the cross-validation DSA

Briefly, we describe some of the key aspects of the theory and methodology behind the cross-validation DSA algorithm in four areas: 1) loss functions as a measurement tool of model competency (e.g., model fit) for fitting the causal parameter of interest in the counterfactual world, and the basis by which these functions are applied to the observed data with the use of available mapping options for counterfactual data (e.g., G-computation, IPTW, DR-IPTW estimators); 2) two methods to estimate the loss function: empirical risk and cross-validation risk; 3) generation of candidate models; and 4) selection of nuisance parameter models. A detailed description of the mechanics of the DSA algorithm is provided in Appendix 1 and 2.

4.1 Loss Functions and Mapping of Counterfactual Data

Loss functions are criteria used in statistics to evaluate and to compare models based on fits of candidate estimators to data (16). With these criteria, and under the assumption that the class of models is an appropriate summary of the data, the goal of selection of a well-fit model occurs by minimization of the expectation of the loss function, or empirical risk, that can be represented generically as

$$\frac{1}{n}\sum_{i=1}^{n} L(O_i, \psi_n(A_i, V_i))$$

with observations O_i on which a candidate model ψ_n is fit. A simple example of a loss function is the squared residual of an observed outcome and the predicted value that has the property that its expectation is minimized by the true conditional mean of the outcome, given the covariates. This loss function is thus suitable for regression. In fact, our nuisance parameter model selection for fitting the conditional mean of the outcome, given exposure and baseline covariates, and for fitting a regression of the exposure on the baseline covariates, is based on this loss function. Another example of a loss function is the minus log likelihood function that has the property that its expectation is minimized by the true density of the data. This is a loss function for model selection for the conditional distribution of a binary outcome, given baseline covariates. These two loss functions are simple in that they are known functions of the data structure and a candidate model for the parameter of interest.

The cross-validation DSA algorithm is based on the estimation of the expectation of loss functions (i.e., risk) where the loss function is a function of the observed data structure $O=(W,A,Y)$ and a candidate MSM for the causal parameter of interest. These loss functions are selected such that their expectation measures the discrepancy between a candidate fit of the causal parameter of interest for different models to the observed data and the true causal parameter (i.e., a perfect fit of a true model to the data).

The true causal parameter is the "absolute" minimizer of the expectation of the loss function (i.e., the risk function). However, in the real world, the true causal parameter is unknown. The goal of selection of the best estimator (i.e. best candidate model) is to find the estimator closest to the true causal parameter. Since the unknown, true causal parameter gives a fixed risk (which is also unknown), the goal is simply to minimize the risk over all candidate model-specific fits of the causal parameter of interest. This can be achieved approximately by minimization of the empirical risk (i.e., the empirical mean of the loss

function).

Another characteristic of loss functions particular to MSMs is the uniqueness of the data structure (i.e., counterfactual as compared with observational data) with which they are computed. If we could observe the counterfactual data X on each subject, we could choose as the loss function for the treatment specific mean (MSM), possibly conditional on some baseline covariates V, the summed (over all possible exposures, a) squared residuals between the counterfactual outcomes under treatment a and a candidate fit of the treatment specific mean**:** $L(X,\psi) = \sum_{a \in A} (Y(a) - \psi(a,V))^2$. This loss function is the standard loss function for repeated measures regression in which each subject has multiple possible outcomes. Indeed, the expectation of this loss function is minimized at the true treatment specific mean of the outcome (i.e., the causal parameter of interest). However, this loss function is not appropriate for the data we typically observe (i.e., single counterfactual outcome that corresponds to a one-time exposure for a given individual).

van der Laan and Robins have presented a method to map counterfactual data-estimating functions to observed data-estimating functions with the same expectation; this method has direct implications for mapping counterfactual data loss functions to loss functions for observed data that can be carried out with any one of the three MSM estimators: G-Computation, IPTW, and DR-IPTW(4, 11). For example, mapping of the loss functions based on the IPTW estimator is formulated as :

$$L_{IPTW} = (Y-\psi(A,V))^2 \, g(A,|V)/g(A,|W)$$

where $g(A|W)$ and $g(A|V)$ are models for the treatment (exposure). The expected values of this IPTW loss function and the loss function for the counterfactual data are equivalent when we assume no unmeasured confounders. The DR-IPTW loss function not only provides a correct model specification for the treatment specific mean (as the IPTW loss function), but it does so with minimum variance (11). Thus, the loss functions that are used typically in conventional analyses to evaluate models can be extended to evaluate and compare MSMs by mapping them into observed data loss function with the G-computation, IPTW or DR-IPTW estimators. The details of the G-computation and DR-IPTW loss functions are provided in Appendix 1.

As stated above, given the model space of potential candidate models is infinite dimensional, selection of these models based on minimization of empirical risk alone would result in over-fit models. Fine tuning parameters that describe the size and complexity of the model needs to be based on so-called cross validation risk. Details of the formulation and application of cross-validation risk as a model selection criterion, in connection with empirical risk, in the cross-validation DSA algorithm are given below.

4.2. Comparison of empirical risk and cross-validation risk used in cross-validation DSA

The algorithm uses a combination of both empirical risk and cross-validation risk criteria that operate jointly to evaluate and select models for the nuisance parameters (appropriate estimators of the MSM---i.e. controls for confounding) and the MSMs themselves. The process by which models are built, compared, and selected, based on

minimization of the empirical risk, is done in the framework of V-fold cross-validation. After models are selected (based on empirical risk) that are representative of the different subspaces (approximate the whole model space) based on training data subsets of the entire data, they are fit with the remainder of the data (correspondent validation data of those training data subsets). The model correspondent with one of the subspaces (e.g., model with 5 terms and no interactions) that minimizes the empirical mean over the different validation datasets—i.e., cross-validation risk average-- is used to select the optimal subspace. The following description elaborates more on this process.

4.2.1. Cross-validation risk

Candidate models of the causal parameter of interest are selected by DSA based on lowest empirical risk estimated from training data for different model size-complexity combinations. These different size-complexity combinations are referred to as subspaces. Since a model of optimal size and complexity cannot be selected based on empirical risk (i.e., model of maximum size-complexity would be selected), the cross-validation-risk is used to select the size and complexity, within bounds set by the user. Cross-validation risk estimates are obtained for these selected models, indexed by size and complexity, based on validation data. For example, a candidate model is estimated with the observations in the training set; then, the empirical mean is taken over the validation set of the loss function at the candidate fit. This consecutive process can be described as a "one-step cross-validation risk" which we define as the empirical mean (average) of the loss function over a validation sample. In the 5-fold cross-validation process, five one-step cross-validation risks are calculated for the different training-validation dataset combinations. A final cross-validation risk is calculated from an average of the five one-step cross-validation risks. It is well known that V-fold cross-validation provides a better estimate of true risk than a single split of the data.

In summary, the cross-validation DSA algorithm maps a training dataset into a set of models that are the optimal models of the subspaces in which these models occur. These models, in turn, are applied to a validation dataset to assess the cross-validation risk for each subspace that each model represents. The process is repeated for V-fold divisions of the whole dataset. At the end, the subspace with the lowest average cross-validation risk is selected as the optimal subspace. The final implementation of the cross-validation DSA algorithm on the whole dataset, provides a set of best models that correspond to each subspace. The final optimal model is the one among these best models that occurs in the optimal subspace (See Figure 1).

5. Generation of candidate estimators with the DSA algorithm

In the DSA algorithm, the whole model space is parameterized as a transformation (e.g., identity or logit function) of linear combinations of basis functions. The choices of the basis functions include the polynomial powers (i.e., 1, x, x^2, …), and spline functions of fixed degree with corresponding fixed set of knot points and wavelets functions. This

choice of a class of basis functions can itself be chosen with cross-validation. The current approach of the DSA is focused on use of the polynomial powers as the basis function.

Given this parameterization, the subspaces are obtained by restrictions on different conditions, i.e., the number of terms that a model contains (k_1), maximum order of interactions (k_2), etc. An example of such a subspace might be models that are polynomial functions of 5 terms and up to 3-way interactions. The best subspace, indexed by $\{k_1, k_2\}$, is selected with V-fold cross-validation. The final model fit is the one that minimizes the empirical risk, based on all the sample data over the subspace selected with V-fold cross-validation.

The minimization in the sequence of subspaces is accomplished with the DSA algorithm. The intuitive idea is that the algorithm searches for a better model in the 'neighborhood' of a 'current best model'—e.g., $A+ AV_1 + V_2$. This 'neighborhood' is defined as the deletion, substitution and addition sets of the current best model. Given the current best model above with $k=3$ terms, the deletion set contains models of size $k-1$ terms, by deletion of one of the k terms from the current model and keeping the other terms (e.g., $AV_1+ V_2$). The substitution set contains models of the same size k (e.g., $A^2+AV_1+V_2$), where each of the k terms is replaced by a new term, respectively. The addition set contains models of size $k+1$, by adding a new term (a single variable or a new term generated by substitution) to the current model (e.g., $A+AV_1+V_2+V_1$).

To evaluate models, the algorithm starts with an intercept model of size $k = 0$. . Then the algorithm performs an addition move, where only a main effect term is added each time (the addition set of the current intercept model only includes main effect terms). The best model of size $k = 1$ is the one that has the minimal empirical risk (mean squared residual) among all the univariate models. The minimal empirical risk and the best model of size $k = 1$ are then saved.

Next, since there is only one term in the model, the algorithm carries out a substitution move. At this point, the algorithm is not interested in the deletion set that returns to the intercept model, and may conduct additional substitutions with interaction terms depending on the number of n-way interactions specified by the user. Within the substitution set, the substitution move finds the minimal empirical risk and its corresponding model and compares this minimal empirical risk with the previously saved minimal empirical risk of size $k = 1$. If the empirical risk of the substitution move is less, the minimal empirical risk and the best model of size $k = 1$ will be updated and a new round of substitution begins. If the empirical risk of the substitution move is not less than the saved empirical risk, the algorithm will keep the previous model and go to an addition move by adding a second term.

Once there is more than one term (exclusive of the intercept) in the model, the algorithm will perform a deletion move first. The deletion move finds the minimal empirical risk and its corresponding model within the deletion set and compares this minimal empirical risk with the previously saved minimal empirical risk of size $k-1$. If the empirical risk of the deletion move is less, the minimal empirical risk and the best model of size k-1 will be updated and the algorithm goes back to a new round of deletion moves (i.e., if there are at least two terms left in the model) or substitution moves (if only one term is left in the model). If the empirical risk of the deletion move is not less than the saved empirical risk of k-1, the algorithm will keep the previous model and go to a substitution

move. Addition moves will be considered up to a maximum model size as specified by the user. The DSA algorithm reports the best model for each size k.

If the subspace also is restricted by maximal order of interactions (e.g., 2-way interaction or 3-way interaction), the deletion, substitution and addition sets are generated under this additional restriction. For example, if the allowed maximal interaction is 3-way interaction, then the DSA algorithm will be carried out three times; first, for models with no interactions; second, for models that include 2-way interactions; finally, for models that include 2- and 3-way interactions. The DSA algorithm returns the best models for all possible combinations of size and level of interaction allowed by the user.

6. Data-adaptive estimation of nuisance parameter models

Since the estimation of MSMs depends on nuisance parameter models, it should be emphasized that the cross-validation DSA needs to be applied to select these models as well, if they are unknown. Selection of a model (i.e., estimator) can occur by a number of different approaches. In each case, the selection of the model precedes the fit of the MSM. One approach to selection of the model (e.g., IPTW estimator) is through an integrated step within the cvMSM component of the cvDSA procedure, described in section 1. Specifically, the training sample that is used to fit a candidate MSM is treated as a whole sample and is split into a number of subsets. Each of these subsets represents training and validation datasets at the level of the fit and selection of the treatment model. Once a candidate treatment model is selected through this process, it is used to estimate the candidate MSM. Other approaches to the selection of the IPTW treatment model, as well as other estimators, are the cvGLM procedure, available with the cvDSA package, or the alternative R-package (DSA) which is used to fit conditional models exclusively. These models would be submitted directly to the cvMSM.

In practice, when one considers treatment models, one examines the association of the outcome variable with each covariate that one considers a potential confounder of the causal effect. Only those variables that are associated with the outcome (e.g., $p<0.2$) are included in the selection of the treatment model.

Different criteria can be used to select treatment models. For example, a model selection criterion was proposed to select the treatment model by minimizing the mean-squared error of the estimator of the MSM (3). However, this joint selection of the treatment and MSM models is not implemented in the current cvDSA R-package. Instead, the criterion used to select treatment models is based on a simpler set of computations which involve minimization of the cross-validated mean-squared error of the treatment model itself.

Ultimately, selection of one of the available estimators depends on which among them can provide consistent MSM estimates. Also, the fit of the MSM will depend on the fit of the nuisance model given the extent of the data to address the ‘no unmeasured confounders’ assumption and the selection criteria provided by the user to the algorithm.

7. Assessment of cross-validation DSA algorithm for model selection

7.1 Overview

Simulations were carried out to assess the performance of the cross-validation DSA for selection of both conditional models and MSMs under a variety of controlled conditions. Data were typically generated based on a random model (e.g., Y~X), where the set of covariates X (e.g., x1-x4) was comprised of random uniform variables, and the parameter values in front of the Xs were generated randomly from the uniform distribution. Random error was incorporated as part of the model as well. The DSA algorithm was utilized to select models closest to the models that generated the data--- i.e., to select the model Y ~ X, for several replicates of data. Bias and mean square error (MSE) were determined for the different DSA-selected models to assess model proficiency for each of these replicates.

A similar approach was taken to assess DSA-selection of MSMs; however, in addition, a binary treatment variable was generated, and the data were simulated to invoke confounding between this treatment variable and the outcome Y. In addition, this latter simulation included comparisons in bias and MSE between MSMs selected by DSA and arbitrary, assumed MSMs, for the different simulated data.

Modifications were made to this overall scheme to assess the cross-validation DSA. Details that pertain to each of the different simulations are given below. Results of the simulations are provided in Tables 1a and 1b, and Figures 2-5 in section 7.3.

7.2 Simulation Methods

Simulation Study I

This simulation evaluated DSA model selection in the presence of varying sample size and random noise (error). The simulated data were based on four covariates x1-x4, all random uniform, and a fixed data-generating model for Y: -1 + x1 + x2 + x1x3 (See Table 1a, Study 1). Datasets of different sizes N (i.e., 500, 1000, 5000) were generated, and random error was added to the outcome Y based on the standard normal distribution with standard deviation σ (i.e., 0.25, 0.5, and 1.0), in separate instances.

The DSA algorithm for selection of conditional models was used, and the selection criteria that were submitted included 5-fold cross validation; a maximum model size=4; orders of interaction=2; and maximum sum of polynomial order or interaction=2. An illustration of how these selection criteria are submitted to the algorithm is provided in Appendix 3.

Bias of each selected model was estimated based on the difference $E(\theta)-E(\theta')$, where θ represents the vector of predicted values based on the data-generating ('true') model and θ' represents the vector of predicted values based on the DSA-selected model. MSE was calculated as Bias^2 + variance, where variance was estimated as $\Sigma(Y-\theta')^2/(n-p)$, where Y represents the vector of observed responses, n is equal to the number of observations, and p is equal to the number of parameters in the DSA-selected model.

Simulation Study II

This particular simulation assessed DSA model selection where random models were used to generate the simulated data. Again, DSA model selection was evaluated given different sample sizes and random error imposed on Y. The simulated data were based on four covariates x1-x4, as in the previous simulation, except the data-generating model for Y was not fixed but random for each simulated dataset (See table 1a, Study 2). Values of the coefficients in the random models were generated from the uniform distribution, and the signs in front of the coefficients (-1, 0, 1) were randomized. Datasets based on different sample sizes (i.e., 500, 1000) were generated, and random error σ (i.e., 0.25, 1) was assigned. The selection criteria submitted to the algorithm were the same as those used in the previous simulation. Bias and MSE were determined for each of the models selected as described above.

Simulation Study III

This next simulation assessed the specificity and sensitivity of DSA model selection. The simulated data were based on additional covariates x1-x10, where some of these covariates were random binary variables. The data-generating model for Y was not fixed but random for each simulated dataset (see Table 1a, Study 3). In addition, the data-generating model consisted of 4 terms or 10 terms, in separate instances, to determine how well the DSA selected a smaller model (specificity) or a larger model (sensitivity) given the additional covariates that were added as part of the simulation. The simulation was evaluated in the context of varying sample sizes (i.e., 500, 1000), with the same random error (i.e., σ=1) assigned to Y. One model selection criterion was modified from the previous simulations by an increase in the maximum model size from 4 to 10. Bias and MSE were calculated as previously described for each of the selected models.

Simulation Study IV

This last simulation examined the performance of the cross-validation DSA to select MSMs. The simulated data were based on a few covariates (x1-x4), and a binary treatment variable whose values (A = 0,1) were assigned based on a linear combination of covariates (i.e. 0.9 + x1 + 0.5x2 – 1.3x1x3), to invoke confounding between A and Y. Random error was incorporated also as part of the treatment assignment. The data-generating model for Y (an MSM) was not fixed but random for each simulated dataset that was generated (see Table 1b below). Each model consisted of the binary treatment variable, and two covariates (x1, x2) to represent V in the MSM. Parameter values of the different model terms were based on values generated at random from the uniform distribution. The simulation was designed so that x1 appeared randomly in the model as either a square term or as part of an interaction with A (i.e., $x1^2$, Ax1, respectively). The cross-validation DSA algorithm was used to choose models closest to ones used to generate data. By way of comparison with the DSA-selected MSMs, we fit fixed, misspecified (i.e., assumed) MSMs to the data which included the treatment variable, and x1 and x2 as singular, 1^{st} order terms only.

All datasets in this simulation were of equal sample size (N=500), and constant random error (σ=0.25) was added to the outcome Y. Different situations characteristic of MSM analyses were incorporated to assess the adaption of the DSA algorithm to these different circumstances. For example, confounding (i.e., 0.9 + 3x1 + 1.5x2 – 1.3x1x3), and random error (σ =1) were increased in separate simulations. Given the additional time required to select MSMs with the DSA, 200, rather than 500 replicates of data employed in the previous simulations, were used to examine the distribution of bias and MSE for the different selected models.

The model selection criteria used as part of cvMSM( ) included 5-fold cross validation; a maximum model size=4; orders of interaction=2; and maximum polynomial order of the different terms=2. The model for A given covariates (e.g., A~X) provided above was submitted to the algorithm, as well as a model A~V, where V=x1, x2. The models were fit with IPTW estimation which included the option for stabilized weights.

7.3 Simulation Results

Tables 1a and 1b contain representative models from the different simulations and illustrate the extent to which models selected by the DSA approximate those that were used to generate the actual data. Measures of bias and mean square error (MSE) summarize the differences based on the models selected by the DSA and the true models of the data. Particular conditions are shown (Tables 1a-1b, far left column) to illustrate how model selection depends on sample size and random variability in the data. Tables 1a and 1b provide single instances of model selection based on single replicates of data, and the correspondent bias/MSE of the selections in question. Figures 2-5, on the other hand, provide distributions of bias and MSE correspondent with models selected by the DSA based on multiple replicates of data, where data were generated repeatedly, and fit with models, for each of the simulations.

Table 1a. Representative DSA selections of conditional models and measures of model proficiency based on simulation studies 1-3.

Study	N	σ	Data-Generating Model	DSA-Selected Model	Bias	MSE
1	500	0.25	$-1 + x_1 + x_2 + x_1x_3$	$-1.04 + 1.05x_1 + 1.04x_2 + 0.91x_1x_3$	0.014	0.055
1	1000	1.00	$-1 + x_1 + x_2 + x_1x_3$	$-0.76 + 0.74x_1^2 + 0.96x_2 + 0.91x_1x_3$	- 0.044	0.999
2	500	0.25	$-9.09 + 0.96x_2$ *	$-9.09 + 0.95x_2 + 0.06x_3$ †	- 0.015	0.060
2	1000	1.00	$-5.65 + 7.04x_1x_3$	$-5.64 + 7.09x_1x_3 - 0.02\ x_1x_2$ †	0.034	1.008
Selection Specificity‡						
3	500	1.00	$8.72 - 8.98x_4 - 1.22x_9$ §	$7.63 - 9.00x_4 - 1.33x_9$	- 0.038	0.945
3	1000	1.00	$5.03 + 2.86\ x_5$	$6.55 + 2.90\ x_5$	- 0.019	0.978
Selection Sensitivity‡						
3	500	1.00	$6.63 - 4.55x_4 - 2.54x_7 + 4.77x_9 - 4.63x_{10}$ §	$7.52 - 4.67x_4 - 2.40x_7 + 4.67x_9 - 4.62x_{10}$	0.056	1.049
3	1000	1.00	$4.36 - 0.56x_1 - 2.40x_8 - 4.32x_9 + 9.09x_{10}$	$4.58 - 0.59x_1 - 2.34x_8 - 4.30x_9 + 9.08x_{10}$	- 0.026	0.999

*Signs for x_1 and x_1x_3 were randomly set to 0; signs could be randomly set to -1, 0, or 1 times the parameter value that was randomly assigned for each term.

† Additional but negligible terms represent some of the non-specific noise which the algorithm specified as part of the model.

‡ Specificity and sensitivity represent algorithm's capacity to select smaller and larger (e.g., 4- and 10-term) models given expanded list of covariates and increased maximum model size to search and select models.

§ Coefficients for the different terms that do not appear in data-generating model were randomly set to 0. Other data-generating models used in the simulation would have included coefficients and terms that do not appear here.

Overall, there was a tendency among selected models toward more bias and variance given more random error in the data (Figure 2, σ increased from 0.25 to 1). This result is not unexpected given it is difficult to select and fit models given variable data in practice. However, the results showed also that increased bias/variance given increased random variability was mitigated by increased sample size.

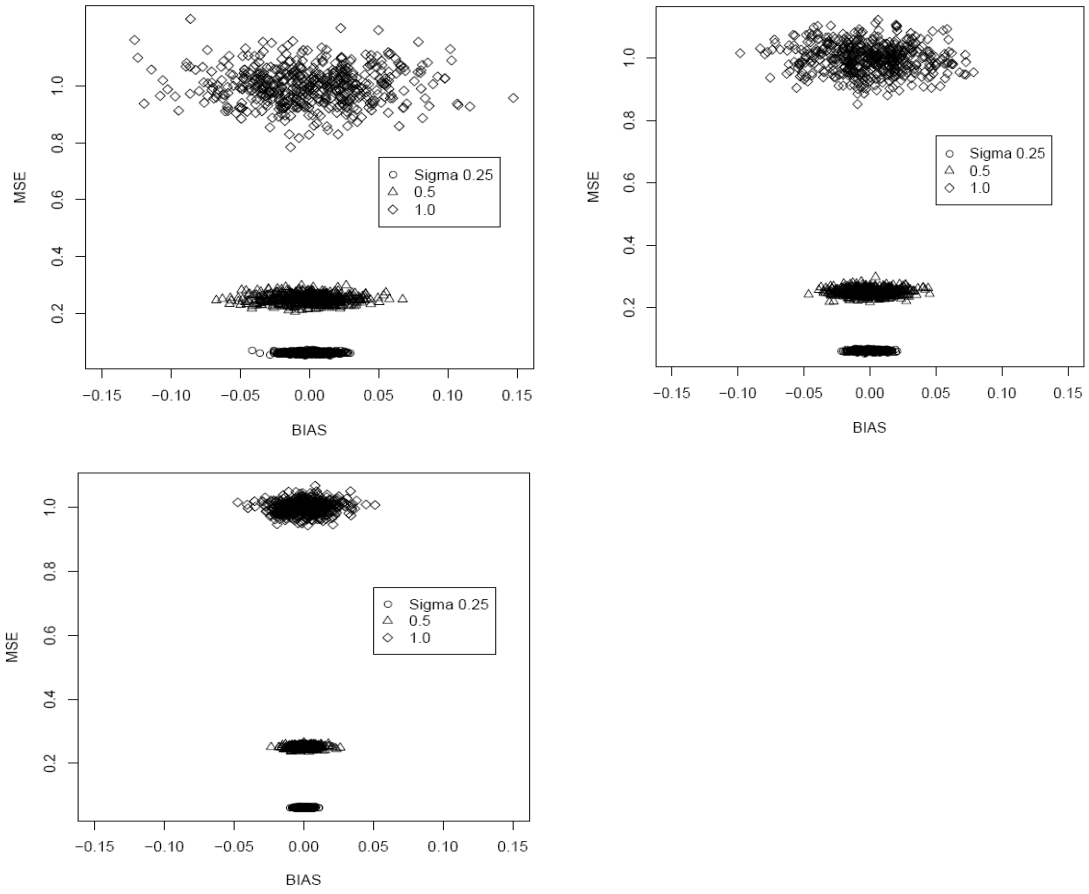


Figure 2. Bias and MSE of DSA-selected models based on 500 replicates of simulated data given varying conditions (sample size: N=500 (top left); 1000 (top right); 5000 (bottom left) and underlying variance σ) where the data-generating model was a fixed model (Simulation Study 1).

The convergence of bias to 0 and the MSE to σ^2, the level of variance in the data which was due to random noise, suggest that the model estimates returned by the DSA are both consistent and efficient.

Comparable results were shown for selected models based on data that were generated based on random models of Y given X (Table 1a-Study 2, Figure 3).

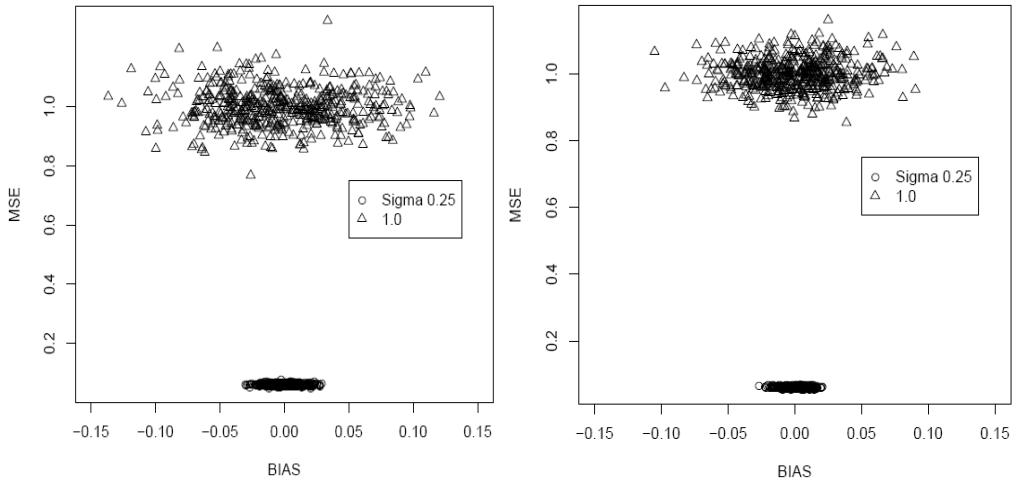


Figure 3. Bias and MSE of DSA-selected models based on 500 replicates of simulated data given varying conditions (sample size: N=500 (left); 1000 (right); and underlying variance σ) where the data-generating model was random for each replicate (Simulation Study 2).

Results were likely similar given: 1) the DSA method of selection was unchanged between simulations 1 and 2; and 2) the random models used in simulation 2 could, in theory, have been represented by the same fixed model used to generate the data in simulation 1.

The DSA algorithm was tested also to determine the specificity/sensitivity with which it selected models (Table 1a-Study 3, Figure 4), based on an enlarged model space--i.e., additional covariates and increased maximum model size.

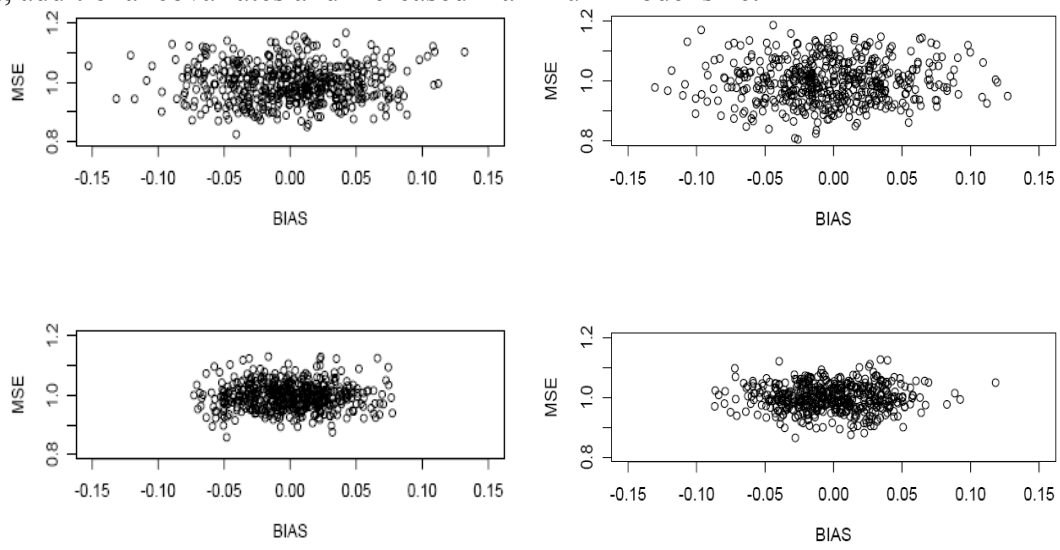


Figure 4. Bias and MSE of DSA-selected models based on 500 replicates of simulated data to examine specificity (left) and sensitivity (right) of DSA-selected models for sample sizes N=500 (top) and 1000 (bottom), given additional candidate variables used in model search (Simulation Study 3).

Given the opportunity to apply several models for given data distributions, the algorithm returned models that approximated the true representative models for a majority of the replicates, whether large or small (See representative models and correspondent bias (Table 1a) to get a sense of relative effect of bias on model results). Moreover, the pattern of bias and MSE of the models in this simulation ---i.e., smaller with increased sample size---was similar to that of previous results.

Results of the simulation with DSA-selected MSMs demonstrated that these models were similar to the true MSMs that were used to simulate the data (Table 1b, Figure 5), however, the results indicated, also, that the models returned by the DSA could be susceptible to additional bias under given conditions.

Table 1b. Representative comparison of DSA-selected and fixed (assumed) MSMs and measures of model proficiency based on simulation study 4 with varying conditions.

Condition*	Model	Bias	MSE
Baseline levels of confounding and random noise†			
True MSM‡	$\boldsymbol{7.56 + 9.68A + 9.10x_2 + 2.87Ax_1}$		
DSA-Selected	$7.56 + 9.67A + 9.12x_2 + 2.85Ax_1$	-0.006	0.064
Fixed MSM	$6.85 + 11.11A + 1.41x_1 + 9.12x_2$	- 0.004	0.242
Increased confounding and baseline random noise§			
True MSM	$\boldsymbol{1.39 + 2.05A + 4.70x_2 + 3.13x_1^2}$		
DSA-Selected	$1.35 + 2.13A + 4.71x_2 + 3.13x_1^2 - 0.23Ax_1$	-0.003	0.055
Fixed MSM	$0.94 + 2.05A + 3.05x_1 + 4.62x_2$	0.014	0.118
Baseline confounding and increased random noise§§			
True MSM	$\boldsymbol{2.01 + 2.22A + 4.06x_2 + 0.35x_1^2}$		
DSA-Selected	$2.00 + 2.02A + 4.26x_2 + 0.46Ax_1$	0.019	1.012
Fixed MSM	$1.85 + 2.22A + 0.35x_1 + 4.27x_2$	0.004	1.010

*Results are based on N=500 with varying conditions (levels) of confounding and random error.
† Baseline levels of confounding represented by g(A|X) =0.9 + x1 + 0.5x2 – 1.3x1x3 and random error σ=0.25 were incorporated in simulated data.
‡ MSM used to generate the simulated data.
§ Increased levels of confounding represented by g(A|X) =0.9 + 3x1 + 1.5x2 – 1.3x1x3 were incorporated in simulated data.
§§ Increased levels of random error defined by σ=1.

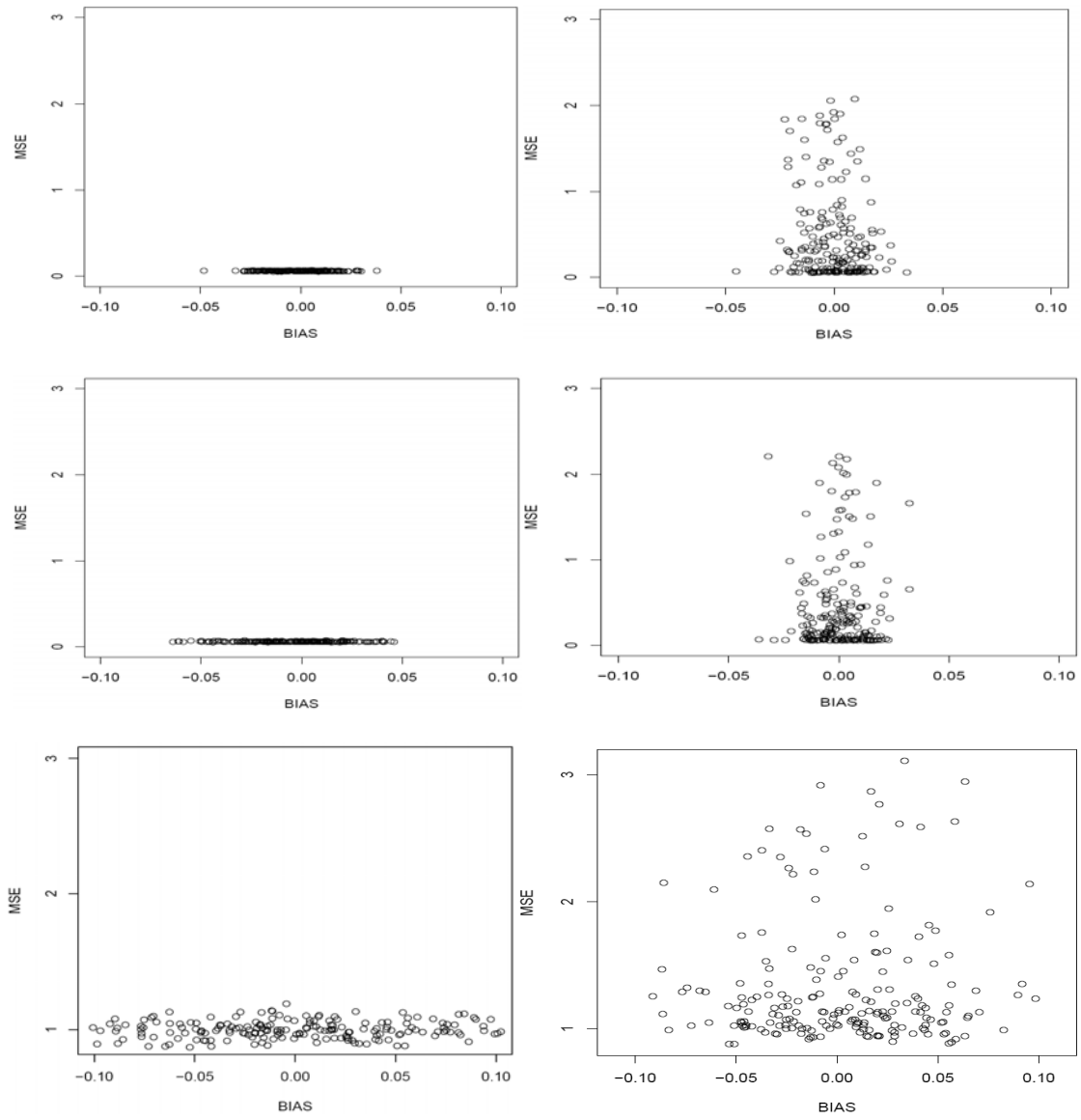


Figure 5. Bias and MSE of DSA-selected MSMs (left) and Fixed (user-specified) MSMs (right) based on 200 replicates of data for varying conditions: baseline levels of confounding and random error (1st row); increased confounding (2nd row); and increased random error (3rd row)).

The results showed that increased random error and confounding, respectively, contributed toward greater bias of the selected MSMs (Figure 5). These sources of bias can be mitigated by increased sample size, in the case of random error, and alternative estimators of MSMs (e.g. G-computation), in the case of confounding. Compared with the fixed (assumed) MSM, the DSA-selected models had more bias but significantly smaller MSE. Based on the selection criteria submitted to the DSA algorithm for MSM selection

(i.e., square terms, 2-way interactions between covariates), the gain in terms of the models was decreased variance but at some cost in bias. Additional bias was observed for the DSA-selected MSMs given greater confounding. The DSA-selected models were based on IPTW estimation, and, for several of the replicates, were biased most likely because of ETA violation. An examination of data for some of the replicates revealed that predicted probabilities of treatment were correlated with observed treatment levels, thus violating a key assumption for identification of causal effects: all treatment levels are observed given covariates (data not show). ETA violation can occur as the result of increased confounding, and is known to lead to biased IPTW estimates (11, 13). Both increased bias and variance were observed for the DSA-selected and fixed MSMs when additional random error was imposed on the data. However, the sample size on which these results were based was 500. The bias and variance of the DSA models, due to random error, would be mitigated with a larger sample.

In summary, the findings from the simulations indicate that the cross-validation DSA is a highly effective tool for model selection, and provides models with consistent estimates and minimal variance. Moreover, the simulation based on MSM selection clearly demonstrated the advantages of the cross-validation DSA for explaining underlying variability that could not be achieved with an assumed model---even in the ideal situation, as represented in this simulation, where the terms of the assumed model were known to be close to those of the true models of the data.

8. Illustrative real-data analysis with the cross-validation DSA algorithm for selection of MSMs

8.1 Overview

The cvDSA algorithm was used to select MSMs to answer the following question: does a population-level 1-liter increase in FEV_1 (forced expiratory volume in 1 second)—a continuous measure of lung function, reduce the hazard of cardiovascular mortality, given age and sex in subjects 55 years and older with no history of active smoking? The objective of the analysis was to use a point treatment study to demonstrate the use of the DSA and cvDSA packages for selection of treatment models and MSMs, respectively, and to compare the models selected by these routines with those models that might otherwise be assumed by an investigator for this type of analysis. The portion of the analysis that involved the selection of treatment models represented the application of the DSA package, given available data, for satisfaction of the no unmeasured confounders assumption---one of the assumptions that is required for the identification of the causal effect of interest.

8.2 Subject characteristics

Data were from a study population of 1053 subjects (716 women, 337 men) with no history of active smoking from a larger longitudinal study of older adults (17, 18), which were examined in a previous analysis (19). One-hundred thirteen cardiovascular deaths occurred in this group, for which the average length of follow-up time was approximately

8.5 years. FEV_1 was measured at the study baseline at ages participants entered the study. Various covariates were collected for the study, and distributions of these are provided in Table 2.

8.3 Model Comparisons

For the purposes of the comparison, the following were examined: 1) an assumed MSM based on an assumed treatment model; 2) assumed MSMs based on DSA-selected treatment models; and 3) DSA-selected MSMs based on DSA-selected treatment models. The assumed MSM throughout the analysis was a Cox proportional hazards MSM:

$$\lambda_{T_{FEV1}}(t \mid Age, Sex) = \lambda_0(t)\exp(\beta_1 FEV_1 + \beta_2 Age + \beta_3 Sex)$$

to evaluate the effect of a population-level 1-liter increase in baseline FEV_1 on the subsequent underlying baseline hazard of cardiovascular mortality $\lambda_0(t)$ over an 8-year period, given age and sex. The model is similar in form to one applied previously (10).

Table 2. Distribution of baseline characteristics and cardiovascular-related mortality in 1053 older adults with no smoking history from the Study of Physical Performance and Age-Related Changes, Sonoma, 1993-2003.

	Females		Males	
	Non-cardiovascular-related mortality	Cardiovascular-related mortality	Non-cardiovascular-related mortality	Cardiovascular-related mortality
N	648	68	292	45
Age[*]	70.0 (8.5)	80.8 (7.8)	68.3 (8.0)	79.9 (6.9)
FEV_1[*†]	2.1 (0.5)	1.4 (0.4)	3.2 (0.7)	2.7 (0.6)
LDL/HDL[*†]	2.5 (1.1)	2.6 (0.8)	3.1 (1.1)	2.8 (0.9)
BMI[*]	26.6 (4.9)	25.4 (4.9)	27.2 (3.7)	26.2 (3.9)
SHS [*,‡]	22.3 (14.2)	26.5 (19.4)	25.8 (13.6)	27.2 (17.8)
SHS [‡,§]	13, 20, 31	12, 25, 40	18, 25, 35	15, 27, 40
Length of follow-up (months) [*]	107.5 (20.9)	61.5 (29.5)	108.0 (18.0)	56.0 (32.9)
Cardiovascular Disease N, (%)				
Yes	59 (9.2)	28 (41.2)	56 (19.4)	17 (37.8)
No	585 (90.8)	40 (58.8)	233 (80.6)	28 (62.2)
Diabetes N, (%)				
Yes	25 (3.9)	8 (11.8)	19 (6.5)	4 (8.9)
No	623 (96.1)	60 (88.2)	273 (93.5)	41 (91.1)

[*] Mean, (SD)
[†] Missing values (not mutually exclusive) Females: 216 (LDL/HDL), 356(FEV_1); Males 72(LDL/HDL), 168(FEV_1)
[‡] SHS (Maximum years from domestic and workplace second-hand smoke exposure)
[§] 25^{th}, 50^{th}, 75^{th} percentiles

8.4 Nuisance parameter "treatment" model selection

Different candidate covariates from table 2 were considered for the treatment models, which were used to derive weights and identify the causal parameters of the various MSMs based on IPTW estimation (9). Given that age, sex, cardiovascular disease, and second-hand smoke (SHS) exposure were associated with the outcome, and potentially associated with FEV_1, these variables were included in an assumed model (see Table 3, Treatment Model I). These variables were submitted to the DSA procedure (see formulation in Appendix 3, Part A) which fit a model with up to 8 terms that could consist of 2^{nd} order polynomial terms and 2-way interactions (Table 3, Treatment Model II). The list of covariates was expanded to include body mass index (BMI) and measures of serum cholesterol (i.e., HDL, LDL) that one might want to consider, although these were not associated with cardiovascular mortality in these data. Based on this adjusted list, the DSA procedure selected a different treatment model (Table3, Treatment Model III). Diabetes, an important cause of cardiovascular disease, was considered as a potential confounder, but was not included in the models that were selected.

8.5 MSM Selection

To select and estimate MSMs, the various treatment models above were specified as parameters in the cvDSA algorithm (see "gaw" in Appendix 3, Part B). Other parameters that were specified as part of the algorithm included: 1) variables that were potential covariates in the various nuisance parameter models ("W"); the baseline covariates---age and sex---included in the MSMs ("V"); and the model formulation based on age and sex ("gav"), used by the algorithm, in conjunction with the specified treatment models, for the development of stabilized IPTW weights.

All MSMs were fit with weighted pooled logistic regression to approximate a Cox proportional hazard regression (20), where each subject contributed data for each 6-month interval that she/he was in the study up to the time of death or loss-to-follow-up. In addition to a fixed IPTW weight calculated for each person by the algorithm, each subject contributed a censoring weight based on the likelihood of being observed for the time she/he was in the study. These censoring weights were developed based on separate models used to estimate each subject's probability of missing FEV_1 and/or serum cholesterol, which were systematically missing variables, and each subject's probability of loss to follow-up. These models were not selected with the algorithm; rather models that included covariates from the treatment models which were significant predictors ($p<0.05$) of censoring were retained. A more formal analysis would have selected censoring models based on the algorithm, since consistent estimation of causal parameters depends on the proper specification of both the treatment and censoring models. The cvDSA algorithm was then used to fit MSMs with up to 6 terms (i.e., a saturated model of FEV_1, age, and sex) that could consist of 2^{nd} order polynomials and 2-way interactions. An added specification partitioned data at the subject-level, rather than at the record-level, for purposes of cross-validation, given repeated observations occurred for each subject.

Standard errors (SE) were calculated for some selected models based on a non-parametric bootstrap (i.e., 1000 samples) (7). For each bootstrap sample, the data were refit with the treatment model and MSM selected in the original data analysis, and the IPTW estimator was recalculated. Thus, the SEs obtained were 'true' to the extent that these models were the 'true' models for each bootstrap sample.

9. Results of data analysis with cross-validation DSA

Representations of assumed models and models selected by the DSA algorithm, for both the treatment models and the MSMs themselves, are given in tables 3 and 4, respectively. The cross-validated (cv) risk estimates listed next to the models represent the associated average "risk" of each of these models, based on size and complexity (i.e., interaction terms), as predictors of different partitions of the data with cross validation. The cv risk represents the criterion by which models are selected by the DSA algorithm, with the lowest cv risk representative of the best possible model given user-defined search criteria (e.g., maximum size, levels of interaction).

In table 3, the differences between the various treatment models that were selected were apparent, given the differences in cv risk and the models themselves. The DSA-selected model that included BMI was the best predictor of the "treatment", FEV_1. Estimated coefficients based on this model indicated that FEV_1 was lower for women than men, and was lower with increased age and BMI, albeit these effects varied with respect to the levels of other variables in the model.

Table 3. Assumed and cross-validation DSA-selected treatment models[*] for FEV_1 in 1053 subjects from the Study of Physical Performance and Age-Related Changes, Sonoma, 1993-2003.

Treatment Model[†]	Type	Model Formula	Cross-Validated Risk[‡]
I	Assumed	$FEV_1 \sim 3.1 - 0.04$ Age $- 1$ Sex + 0.008 CVD $-$ 0.002 SHS	0.201501
II	DSA	$FEV_1 \sim 3.02 - 0.06$ Age $- 0.96$ Sex + 0.02 Age*Sex	0.194245
III	DSA	$FEV_1 \sim 3.52 - 0.06$ Age $- 1.76$ Sex $-$ 0.0007 Bmi^2 + 0.02 Age*Sex + 0.03 Bmi*Sex	0.184641

[*]Selection criterion submitted to the DSA algorithm allowed the procedure to select models up to 8 terms, with 2nd order polynomials and 2-way interactions.
[†]Model II was based on candidate covariates that included age, sex, cardiovascular disease (CVD) and second-hand smoke exposure (SHS) Model III was based on an expanded list of candidate covariates that included body mass index (BMI) and measures of serum cholesterol (HDL, LDL).
[‡]Results can vary if v-fold splits used in the estimation procedure of the DSA are not fixed.

Table 4. Assumed and cross-validation DSA-selected MSMs[*] for the causal effects of FEV_1, age, gender on the hazard of cardiovascular mortality in 1053 subjects from the Study of Physical Performance and Age-Related Changes, Sonoma, 1993-2003.

MSM Model[†]	Type	Model Formula	Cross-Validated Risk
1	Assumed	Intercept Only	0.0127950
2	Assumed	$-5.13 - 0.18FEV_1 + 0.15Age - 1.10Sex$	0.0126576
3	Assumed	Intercept Only	0.0147159
4	Assumed	$-4.55 - 0.33\ FEV_1 + 0.14\ Age - 1.35\ Sex$	0.0144916
5	DSA	$-6.07 + 0.08\ FEV_1*Age$	0.0144237
6	Assumed	Intercept only	0.0119989
7	Assumed	$-4.59 - 0.35FEV_1 + 0.14Age - 1.25\ Sex$	0.0119274
8	DSA	$-6.13 + 0.144\ Age$	0.0119013

[*]Selection criterion submitted to the cvDSA algorithm allowed the procedure to select models up to 6 terms with 2-way interactions.
[†]MSMs based on the application of different treatment models for FEV_1 (see previous table): Models 1-2 (Assumed Treatment Model I); Models 3-5 (DSA-selected Treatment Model II); Models 6-8 (DSA-selected Treatment Model III)

The assumed MSMs reflect the hypothesis that differences in FEV_1 have an overall population effect on the hazard of cardiovascular mortality. However, the smaller models selected by the DSA indicated otherwise. In particular, the MSM selected by the DSA (Table 4, MSM Model 8), based on the best IPTW estimator of the data (Table 3, Treatment Model III), suggested that age alone provided a sufficient fit of the data. A comparison of the results based on the optimal IPTW estimator (Treatment Model III) suggested that a model with age alone fit the data best (Age: β (SE) = 0.14 (0.02); cv risk=0.0119013), followed by a model with age and sex, which was estimated with less precision (Age: β (SE) = 0.16 (0.02); Sex: -0.92 (0.37); cv risk=0.0119224). By comparison, the assumed MSM which included FEV_1 was fit with even less precision (FEV_1: β (SE) = -0.35 (0.43); Age: 0.14 (0.03); Sex: -1.25 (0.44); cv risk=0.0119274).

Plots of cv risk estimates against model size and complexity provide a graphical representation of the relative differences of the models considered in the DSA MSM selection process (Figure 6). For example, the two points with the lowest cv risk estimates in Figure 6 (right side) are representative of the DSA models where FEV_1 was excluded as a main effect.

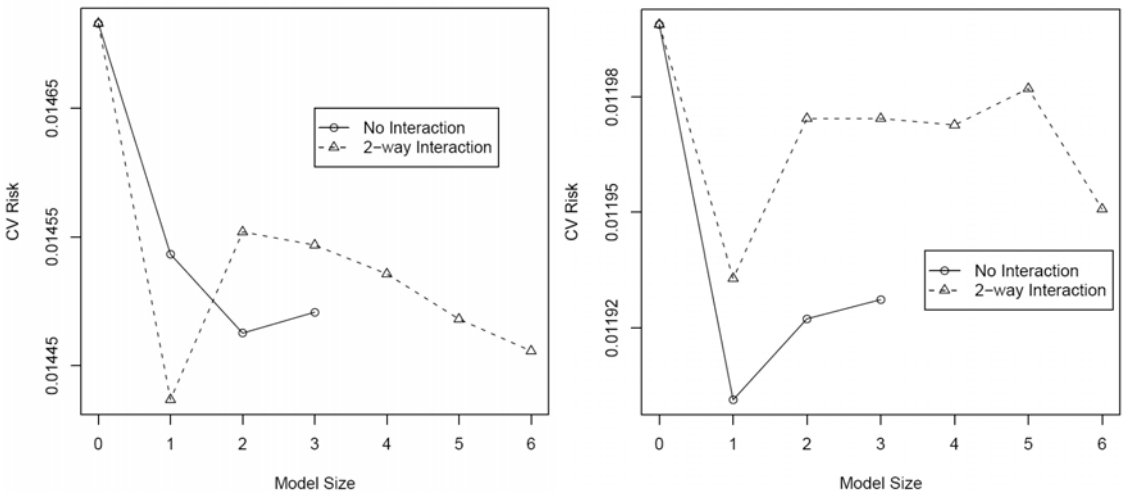


Figure 6. Comparison of Cross-validation (CV) risk, model size, and model complexity (i.e., interactions) based on DSA-selected MSMs #5 (left) and #8 (right).

Separate results compare estimates of an assumed MSM based on an assumed treatment model (Table 4, model 2) with other assumed MSMs that were based on DSA-selected treatment models (Table 4, models 4 and 7). The differences in the results are indicative of the sensitivity of MSM estimates to the choice of treatment models.

In summary, this analysis demonstrates the utility of the DSA algorithm for selection of treatment models and MSMs that would not likely be considered as potential models in practice. Moreover, it highlights the importance of selection of appropriate treatment models for proper estimation of MSM causal parameters.

10. Discussion

The cross-validation DSA algorithm is one of the first model selection procedures written to identify and to estimate causal models for given data distributions and represents an important advancement in the application of MSMs for epidemiologic research. Development of the algorithm represents the combination of: 1) a set of theoretical results that showed that, with cross-validation, an intensive data-adaptive model search can be conducted with finite sample data and that a model, closest in approximation of a true model of the data, can be selected from among many candidate models that might be considered (5, 21); and 2) the development of the Deletion/Substitution/Addition DSA algorithm used to generate and select models, and, thus, approximate the model space for a given data distribution, based on cross-validation (6).

The performance of the algorithm to select models relative to true models of given data was examined with simulations. The algorithm returned consistent models with minimal variance, and returned better representative models of the underlying data, than the fixed MSM models that were evaluated. This finding clearly points to the advantages and practical uses of the algorithm with regard to model interpretability and precision. Still, there is the potential for bias, as suggested by the simulation that included increased levels of confounding. Biased IPTW estimates in particular can occur as the result of large subject specific weights, and reflect a violation of the ETA assumption necessary for the identification of causal effects (22). Different methods can be employed to address this assumption (1, 10, 11, 14, 23). Other estimators are available, too, that can provide consistent and potentially more efficient estimates than IPTW (van der Laan and Robins 2002; van der Laan and Rubin 2006; Robins et al. 2007;Tan 2007; Cao et al. 2009; Goetgeluk et al. 2009).

The real-data analysis provided the opportunity to compare models the algorithm selected with hypothetical assumed models --i.e., *a priori* models. The treatment model selected by the cross-validation DSA demonstrated the algorithm's capacity to identify variables and relationships between those variables that were not originally assumed. Similarly, the algorithm's selection of a MSM that excluded FEV_1 as a causal effect demonstrated the algorithm's choice of a model that was more representative of the underlying data--i.e., FEV_1 in the assumed model was measured with imprecision, therefore there was no clear evidence that it represented a causal effect with the given data. Although the analysis was oversimplified, it illustrated the use of algorithm to examine and recast the modeling assumptions that are applied in data analyses.

The cross-validation DSA represents an important methodological advancement with important statistical and subject matter implications for model selection and MSM analyses. It provides for exploration of a wide assortment of possible models, beyond what current forward/selection procedures can explore, and from these obtain a model closest to the true model of the given data. Consequently, one can have greater confidence in the inferences he/she derives from analyses, given that the models selected do not depend entirely on *a priori* assumptions (they still depend on chosen variables used in the selection process), but are more likely to represent underlying patterns in data. The algorithm is intended as a tool to augment the search of potential causal mechanisms, but is not expected to replace one's discretion in terms of her/ his knowledge of potential underlying causal mechanisms.

The DSA algorithm has additional functions that were not developed for the cvDSA procedure: 1) selection of models based on different numbers of observations, depending on the number of terms with missing values included in the model search; 2) random partitioning of data into training and validation subsets for cross-validation, rather than generating fixed partitions only; and 3) extension of the machine-learning approach where user-supplied models are assessed with respect to model fit, and if necessary, augmented by the algorithm to provide more reasonable fits of given data.

Both the DSA and cvDSA packages are based on an integrated data-adaptive estimation procedure, which enable model searches that are concurrently intensive and robust.

In summary, this paper was intended to illustrate the motivation behind the development of the cross-validation DSA algorithm, examine the mechanisms by which the algorithm selects models, and explore various aspects of the algorithm through simulation and data analysis to inform the researcher who decides to include it among his/her analytical tools.

References

1. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550-60.
2. Mortimer KM, Neugebauer R, van der Laan M, Tager IB. An application of model-fitting procedures for marginal structural models. Am J Epidemiol 2005;162:382-8.
3. Brookhart MA, Laan MJvd. A semiparametric model selection criterion with applications to the marginal structural model. Computational Statistics and Data Analysis 2006;50:475-498.
4. Wang Y, Bembom O, van der Laan M. Data Adaptive Estimation of the Treatment Specific Mean. Berkeley Electronic Press/UCB Biostatistics Division 2004;159.
5. van der Laan M, Dudoit S. Unified Cross-Validation Methodology For Selection Among Estimators and a General Cross-Validated Adaptive Epsilon-Net Estimator: Finite Sample Oracle Inequalities and Examples. Berkeley Electronic Press/UCB Biostatistics Division 2003;130.
6. Sinisi S, van der Laan M. Loss-Based Cross-Validated Deletion/Substitution/Addition Algorithms in Estimation. Berkeley Electronic Press/UCB Biostatistics Division 2004;143.
7. Efron B, Tibshirani RJ. An Introduction to the Bootstrap: Chapman and Hall, 1993.
8. Dudoit S, van der Laan M, Keles S, Molinaro A, Sinisi S, Teng S. Loss-Based Estimation with Cross-Validation: Applications to Microarray Data Analysis and Motif Finding. Berkeley Electronic Press/UCB Biostatistics Division 2003;137.
9. Robins JM. Association, causation, and marginal structural models. Synthese 1999;121:151-79.
10. Hernan M, Brumback B, Robins J. Marginal Structural Models to Estimate Causal Effect of Zidovudine on the Survival of HIV-Positive Men. Epidemiology 2000;11:561-570.
11. van der Laan MJ, Robins J. Unified Methods for Censored Longitudinal Data and Causality, 2002.
12. Yu Z, van der Laan M. Construction of Counterfactuals and the G-computation Formula. Berkeley Electronic Press/UCB Biostatistics Division 2002;122.
13. Neugebauer R, van der Laan M. Locally Efficient Estimation of Nonparametric Causal Effects on Mean Outcomes in Longitudinal Studies. Berkeley Electronic Press/UCB Biostatistics Division 2003;134.

14. Haight T, Neugebauer R, Tager I, van der Laan M. Comparison of Inverse Probability of Treatment Weighted Estimator and a Naive Estimator. Berkeley Electronic Press/UCB Biostatistics Division 2003;140.
15. Bryan J, Yu Z, Van Der Laan MJ. Analysis of longitudinal marginal structural models. Biostatistics 2004;5:361-80.
16. Bickel P, Doksum K. Mathematical Statistics-Basic Ideas and Selected Topics. 2001.
17. Satariano WA, Smith J, Swanson A, Tager IB. A census-based design for the recruitment of a community sample of older residents: Efficacy and costs. Annals of Epidemiology 1998;8:278-82.
18. Tager IB, Swanson A, Satariano WA. Reliability of physical performance and self-reported functional measures in the elderly. Journal of Gerontology: Medical Sciences 1998;53A:M295-M300.
19. Eisner M, Wang Y, Haight T, Balmes J, Hammond K, Tager I. Second-hand smoke exposure, pulmonary function, and cardiovascular mortality. Annals of Epidemiology 2007;17:364-373.
20. D'Agostino RB, Lee M-L, Belanger AJ, Cupples LA, Anderson K, W.B. L. Relation of pooled logistic regression to time dependent Cox regression analysis: The Framingham heart study. Statistics in Medicine 1990;9:1501-15.
21. van der Laan M, Dudoit S, van der Vaart AW. The Cross-Validated Adaptive Epsilon-Net Estimator. Berkeley Electronic Press/UCB Biostatistics Division 2004;142.
22. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol 2008;168:656-64.
23. Bembom O, van der Laan MJ. A practical illustration of the importance of realistic individualized treatment rules in causal inference. Electron J Stat 2007;1:574-596.

Appendix 2

Effects of Body Composition and Leisure-time Physical Activity of Physical Function

Questions Used to Assess Self-Reported Functional Limitation

In the past month, what level of difficulty have you had:

In pushing objects like a living room chair?
In stooping, crouching or kneeling?
In getting up from a stooping, crouching or kneeling position?
In lifting or carrying items under 10 pounds (4.54 kg), like a bag of potatoes
In lifting or carrying items over 10 pounds, like a bag of groceries?
In standing in place for 15 minutes or longer?
In sitting for long periods, say 1 hour?
In standing up after sitting in a chair?
In walking alone up and down a flight of stairs?
In walking two to three neighborhood blocks?

Response categories were:

A lot of difficulty
Some difficulty
A little difficulty
No difficulty
Don't do on a doctor's orders
Don't do because unable
Never do the activity

Exposure History defined in the MSM analysis

In the women's analysis, we define joint exposure history $\bar{a}$ as a function ($\overline{L/F}\,(t) + LTPA\,(t) + \overline{LTPA}\,(t\text{-}1)$) where $\overline{L/F}\,(t)$ represents the "running" mean of L/F through time *t*, $LTPA\,(t)$ consists of 4 levels of exercise pattern at time *t (categorical form),* and $\overline{LTPA}\,(t\text{-}1)$ consists of 3 levels (collapsed 1st and 2nd levels) based on a categorization of the "running" mean of LTPA (continuous form) through time *(t-1)* . In the men's analysis, we define $\bar{a}$ as a function $(\overline{L/F}\,(t) + \overline{LTPA}\,(t))$, where $\overline{L/F}\,(t)$ has a similar representation as it does for women, and $\overline{LTPA}\,(t)$ consists of 3 levels of exercise (collapsed 1st and 2nd levels) based on a categorization of the "running" mean of LTPA through time *(t)* These functions were chosen based on a model selection process that compared different MSMs using different functions of L/F and LTPA history. $\overline{Y}\,(t\text{-}1)$

represents a vector of separate terms of previous functional history that were included in the models, e.g. *Y (1)* for the model at t=*2*; *(Y (1),Y (2))* at t=*3*,and *(Y (1),Y (2),Y (3))* at t=*4*.

Computation of Counterfactual Transition Distributions

We illustrate the method by which we compute the distribution of transitional patterns $\Pr(\overline{Y_{\bar{a}}} = \bar{y})$ for a counterfactual exposure history $\bar{a}$. At each *t* (*t=1,...4)*, we estimate $\Pr(Y_{\bar{a}}(t) = 1 \mid \overline{Y_{\bar{a}}}(t-1))$ with the parameter estimates from the following model at each *t*,

$$\textit{Logit} \quad \Pr(Y_{\bar{a}}(t) = 1 \mid \overline{Y}_{\bar{a}}(t-1)) = \hat{\beta}_{0t} + \hat{\beta}_{1t} * \overline{a}(t) + \hat{\beta}_{2t} * \overline{y}(t-1)$$

where $\bar{a}(t)$ represents exposure history described in Appendix 2; $\hat{\beta}_{1t}$ corresponds to a vector of beta coefficients associated with the components of $\bar{a}(t)$; $\bar{y}(t-1)$ represents past functional history to t, which is a vector that cumulates individual terms of past functional limitation status over time, and $\hat{\beta}_{2t}$ represents the vector of beta coefficients corresponding to those terms of past functional limitation.

We estimate the marginal probability of functional limitation=1 at t=1

$$\Pr(Y_{\bar{a}}(1) = 1) = \exp it\left(\hat{\beta}_{01} + \hat{\beta}_{11}\right)$$

and its complement, the probability of functional limitation=0 as

$$1 - \Pr(Y_{\bar{a}}(1) = 1) = 1 - \exp it\left(\hat{\beta}_{01} + \hat{\beta}_{11}\right)$$

We estimate the marginal probability of functional limitation =1 at t=2 within strata of past functional history as

$$\Pr(Y_{\bar{a}}(2) = 1 \mid \overline{Y_{\bar{a}}}(1) = 0) = \exp it\left(\hat{\beta}_{02} + \hat{\beta}_{12}\right)$$

$$\Pr(Y_{\bar{a}}(2) = 1 \mid \overline{Y_{\bar{a}}}(1) = 1) = \exp it\left(\hat{\beta}_{02} + \hat{\beta}_{12} + \hat{\beta}_{22}\right)$$

and the marginal probability of functional limitation=0 as

$$1 - \Pr(Y_{\bar{a}}(2) = 1 \mid \overline{Y_{\bar{a}}}(1) = 0) = 1 - \exp it\left(\hat{\beta}_{02} + \hat{\beta}_{12}\right)$$

$$1 - \Pr(Y_{\bar{a}}(2) = 1 \mid \overline{Y_{\bar{a}}}(1) = 1) = 1 - \exp it\left(\hat{\beta}_{02} + \hat{\beta}_{12} + \hat{\beta}_{22}\right)$$

(Note, for convenience, we use the same fitted parameter estimates above $\hat{\beta}_{02}, \hat{\beta}_{12}$ to illustrate two different models $\Pr(Y_{\bar{a}}(2) = 1 \mid \overline{Y_{\bar{a}}}(1) = 0)$ and $\Pr(Y_{\bar{a}}(2) = 1 \mid \overline{Y_{\bar{a}}}(1) = 1)$; the parameter estimates $\hat{\beta}_{02}, \hat{\beta}_{12}$ indeed are different each time we add terms to the models. A similar pattern occurs below for *t*=3 and *t*=4.)

For t=3, the marginal probability of functional limitation=1 (marginal probability of functional limitation=0 not shown) within strata of functional history is given by

$$\Pr(Y_{\bar{a}}(3) = 1 \mid \overline{Y_{\bar{a}}}(2) = (y(1) = 0, y(2) = 0)) = \exp it\left(\hat{\beta}_{03} + \hat{\beta}_{13}\right)$$

$$\Pr(Y_{\bar{a}}(3) = 1 \mid \overline{Y_{\bar{a}}}(2) = (y(1) = 1, y(2) = 0)) = \exp it\left(\hat{\beta}_{03} + \hat{\beta}_{13} + \hat{\beta}_{213}\right)$$

$$\Pr(Y_{\bar{a}}(3) = 1 \mid \overline{Y_{\bar{a}}}(2) = (y(1) = 0, y(2) = 1)) = \exp it\left(\hat{\beta}_{03} + \hat{\beta}_{13} + \hat{\beta}_{223}\right)$$

$$\Pr(Y_{\bar{a}}(3) = 1 \mid \overline{Y_{\bar{a}}}(2) = (y(1) = 1, y(2) = 1)) = \exp it\left(\hat{\beta}_{03} + \hat{\beta}_{13} + \hat{\beta}_{213} + \hat{\beta}_{223}\right)$$

Equations for estimation of the probability of functional limitation=1 for t=4 build on the pattern for t=3, except 8 equations are estimated for each of the 8 strata of past functional history.

We can now construct the joint probability of functional limitation over time using the cumulative product of computed probabilities from above

$$\prod_{t=1}^{4} \Pr(Y_{\bar{a}}(t) = y(t) \mid \overline{Y_{\bar{a}}}(t-1) = \bar{y}(t-1))$$

to obtain the marginal distribution of transitions for exposure history $\bar{a}$ for the entire duration of the study $\Pr(\overline{Y_{\bar{a}}} = \bar{y})$.

Appendix 3

Secondhand Smoke Exposure and Cardiovasucular Disease on Dementia Risk

Implementation of Cox Proportional Hazards Marginal Structural Models (Cox PH MSMs)

I. Overview

Marginal structural models (MSMs) are based on the concept of counterfactuals, which represents the set of outcomes subjects might have experienced if they experienced exposures other than the ones actually received(1-4). Hypothetically, if we knew the outcomes that corresponded to all possible exposures that a given subject could experience (i.e., the outcomes associated with their 'actual' experiences as well as their 'counterfactual' experiences), then each subject would serve as their own control and we could assess whether differences in the outcome were attributable causally to differences in the level of exposure. In practice, we do not observe all possible outcomes. MSMs represent one class of causal statistical models for modeling this hypothetical world, based on observed data, and for examining causal parameters (relationships) of interest (2). Moreover, estimation of MSMs account for time-dependent confounders and informative censoring, and therefore can provide unbiased estimates of these causal parameters (1).

Identification of causal effects with MSMs, based on observed data, depend on a set of assumptions that are followed by estimation procedures (e.g. Inverse Probability of Treatment Weight, IPTW) to estimate the causal parameters defined in the models (1). In addition to IPTW weights, weights are obtained to account for systematically missing covariates and loss to follow-up (Inverse Probability of Censoring Weights, IPCW) (1). A weighted logistic regression estimated with generalized estimating equations (GEE), with weights derived for each subject, was used to fit the Cox Proportional Hazards (PH) MSMs in this analysis(5).

Typically, MSMs include both "causal parameters" and "stratification variables." In our analyses, it was not possible to define SHS as the causal parameter because this would have required knowledge of both when SHS exposure occurred and what factors contributed to greater or lesser exposure to SHS (both required for definition of causal effects). Therefore, we examined SHS as a baseline stratification variable and treated clinical cardiovascular disease (CVD) as the casual effect parameter in the analyses. This seemed reasonable, since SHS exposure has been associated with increased risk of cardiovascular disease. Additional baseline stratification variables included subclinical measures of carotid artery disease and cerebral MRI status, age, gender, and education.

This approach enabled us to examine the causal contribution of population-level differences in CVD incidence to the change in hazard of onset dementia at time *t* during the 6-year study. In addition, we examined the associated change in the hazard of dementia for different baseline stratification variables given that, contrary to fact, no one in the population had clinical CVD. In this manner, we examined the association of the different

levels of these stratification variables with respect to the risk of onset dementia independent of these variables' influence on clinical CVD. Identification and estimation of parameters, used to represent the contributions of these variables to the risk of dementia, are based on a set of assumptions that underlie MSMs. For example, confounders (see "Other measures" section of Methods) of clinical CVD and dementia were examined in a separate "treatment" model to satisfy one of these assumptions, and results from that model were applied toward the estimation of the Cox PH MSM parameters. In this manner, one includes only those variables of direct interest in the MSM.

In summary, estimates of Cox PH MSMs represent, more directly, estimates of the causal relative hazard of exposure and disease onset than do the estimates from the usual Cox PH association model. Moreover, these MSM estimates provide unbiased estimates in the presence of causal intermediates that cannot occur with standard analytical methods.

II. Causal Model

Given a time-dependent process *(t=1,...,6)* that includes clinical cardiovascular disease (CVD) incidence, recorded covariates (time-dependent and time-independent), and occurrence of dementia, we define the following Cox PH MSM to examine the causal contributions of population-level CVD incidence and baseline risk covariates "*v*" (e.g., age, cumulative lifetime SHS exposure) to the hazard of onset dementia at time *t*.

$$\lambda_{T_{\bar{a}}}(t \mid v) = \lambda_0(t)\exp(\beta_1 \bar{a}(t) + \beta_2 v)$$

$T_{\bar{a}}$ denotes a counterfactual outcome process of time to dementia (years after baseline interview) that corresponds with all possible courses of CVD incidence subjects could have experienced during the 6-year study $\bar{a}(t)$. Once a subject experienced an event, s/he was classified as having CVD for the length of the study. $\lambda_0(t)$ represents the unspecified baseline hazard of dementia at *t*. The parameter β_1 signifies the causal log relative change to the baseline hazard of dementia at time *t*, if, contrary to fact, everyone in the population experienced clinical CVD through observation time *t*, compared to no one in the population having CVD. These counterfactual hazards were examined for subpopulations defined by *v* (*v* includes categories of exposure to SHS). β_2 is a parameter vector that signifies the associated log relative hazard change to the baseline hazard at anytime *t* for different subpopulations of subjects as defined by *v*, --e.g., category of SHS, age----given everyone in the population did not experience clinical CVD. It is important to note that the parameters that correspond with the baseline covariates are not causal but <u>association</u> parameters; however, these can be interpreted as having taken into account clinical CVD as a causal intermediate.

III. MSM Assumptions

In general terms, we make the following assumptions to identify causal effects with MSMs: 1) temporal ordering of variables---e.g., covariates at *t* precede clinical cardiovascular disease status at *t* and dementia status at *t+1*; 2) consistency assumption—observed data are just one realization of the "full" counterfactual data; 3) no unmeasured confounding – i.e. that the treatment (e.g., clinical CVD) at any given time *t*, conditional on covariates, is independent of the counterfactual outcome; 4) experimental treatment assignment (ETA) which states that all possible treatments (e.g., occurrence/absence of clinical CVD) are observed for given covariates (1-4).

IV. Estimation of Causal Parameters

Identification of Cox PH MSM parameters requires satisfaction of the above assumptions, and estimation procedures to quantify the parameters of interest. Estimation was based on the solution of the inverse probability of treatment weight (IPTW) estimating equation(1, 5). Solution of this equation is equivalent to performance of a weighted Cox PH regression with subject-specific weights:

$$Sw(t) = \prod_{k=1}^{t} \frac{\Pr(A(k) \mid \overline{A}(k-1) = 0, V)}{\Pr(A(k) \mid \overline{A}(k-1) = 0, \overline{L}(k))}$$

These weights were obtained by two probability models. One model was a pooled logistic regression in those with no prior clinical CVD $\overline{A}(k-1)$ of incident clinical CVD $A(K)$ on time-dependent covariates $L(k)$ (e.g. hypertension, physical activity, etc. at time *k*), and time-independent covariates (e.g., gender, education, income), described in the methods section labeled "Other measures". The second was a similar model to first, but was a fit of incident clinical CVD on subsets of the baseline covariates V ($V \subset L(1)$) that were specified in the different Cox PH MSMs. The probabilities from the two models were used to create subject-specific stabilized weights at each time point for which a subject's data were observed.

Informally, the denominator of the formulation above represents the probability of observing a subject's history of incident clinical CVD given covariates and no previous history of clinical CVD through time *k* (5).

In addition to IPTW weights, subject-specific weights were obtained to account for systematically missing covariates and loss to follow-up (Inverse Probability of Censoring Weights, IPCW) (5):

$$Swc(t) = \left(\prod_{k=2}^{t} \frac{1}{\Pr(C(k) = 0 \mid C(k-1) = 0, \overline{L}(k-1))} \right) * \frac{1}{\Pr(C(1) = 0 \mid L^{*}(1))}$$

Here a pooled logistic model was fit to evaluate censoring at follow-up $C(k)$ based on previous levels of covariates $\overline{L}(k-1)$ among those present at time *k-1*. Past clinical CVD

was not considered in the model as a covariate because of limited data. Separate logistic models were used to evaluate systematically missing data in two baseline variables (ApoE, Income) using other baseline covariates $L^*(1)$, and the estimates were used to determine the joint probability of no censoring in either of these variables, $C(1) = 0$. Other missing data at baseline were considered to be missing at random. It was not necessary to develop a set of stabilized weights for these IPCW weights given these were relatively stable (i.e., no large weights). Informally, these weights represent the inverse probability of observing a subject in the data given her/his past covariate history.

We implemented the Cox PH MSM with a weighted logistic regression, for example, of clinical dementia $Y(t)$ on clinical CVD $\bar{a}(t-1)$ and baseline covariates V among those with no previous clinical dementia,

$$\log it \Pr(Y(t) = 1 \mid Y(t-1) = 0, V) = \beta_0 + \beta_1 \bar{a}(t-1) + \beta_2 V$$

with the product of the weights $Sw(t) * Swc(t)$ derived for each subject. Generalized estimating equations (GEE) were used to account for within-subject correlation induced by the use of the weights (5). Given the standard errors with this approach tend to be too conservative, we obtained empirical estimates of the standard errors with 1000 bootstrap replications(2).

V. ETA assumption

ETA violation was assessed informally by comparison of the distributions of the conditional probabilities of clinical CVD given covariates, in subjects with and without CVD. Given the low frequency of clinical CVD in the population, both distributions indicated there was a low probability of disease, but there was sufficient variability in both distributions (i.e., covariates not deterministic of CVD status) to indicate that a violation of the ETA was unlikely.

References

1. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550-60.
2. van der Laan MJ, Robins J. Unified Methods for Censored Longitudinal Data and Causality, 2002.
3. Robins JM. Association, causation, and marginal structural models. Synthese 1999;121:151-79.
4. Greenland S, Robins J, Pearl J. Confounding and Collapsability in Causal Inference. Statistical Science 1999;14:29-46.
5. Hernan M, Brumback B, Robins J. Marginal Structural Models to Estimate Causal Effect of Zidovudine on the Survival of HIV-Positive Men. Epidemiology 2000;11:561-570.

Appendix 4

Direct effects of leisure-time physical activity on walking speed

Targeted Maximum Likelihood Estimation (TMLE)

FX Models. For each evaluation period, *FX* models for men and women were developed that included LTPA, Lnfat, diabetes and other covariates associated with walking speed (See earlier discussion of TMLE in Chapter 2). Cross-validation DSA was used to select models based on a candidate covariate list and other selection criteria (i.e., fitted models could include up to 15 terms, with 2-way interactions between variables and 2^{nd} order polynomial terms). All models included covariates associated with walking speed as candidate covariates---e.g., age, BMI, cardiovascular disease, other health-related problems, depression, socio-economic variables (i.e., education, income, living arrangements). Models based on follow-up evaluations *(t=2,3,4)* included past variables *(t-1)* that were associated potentially with walking speed at *t*—i.e., past physical function, past Lnfat, past LTPA. In effect, the *FX* model selected for any given time point represents the joint effects of the variables in the model on walking speed for that evaluation period. In order to target the effects of LTPA and Lnfat on walking speed, in particular, additional steps were required.

Targeted estimation of Fx Model Parameters. Given the possibility that differences in walking speed due to effects of LTPA or Lnfat, based on the *FX* model may not have been strictly related to these variables---i.e., the *FX* model may not have captured fully the potential confounding due to these variables with respect to walking speed---separate models were constructed for the 2 variables of interest. Each of these models was similar to the 'treatment' model of the form $g(A_{ij}|W_{ij})$ described in the General Methods Chapter, where A_i represents the i^{th} variable (e.g., LTPA) at the j^{th} time point. The set of covariates W_i would vary depending on the i^{th} variable—i.e., it was assumed that LTPA affected Lnfat; therefore, the treatment model for LTPA at time *j* excluded Lnfat measured at time *j*.

Next, based on the estimated $g(A_{ij}|W_{ij})$, each subject received a value $(1/g(A_{ij}|W_{ij}))$ if he/she has a value of A=1; or $-1/1-g(A_{ij}|W_{ij})$ if A=0. These values are included as covariates *h(A|W)* in the following model for the given targeted parameters of interest *i*, separately for each time point *j*:

$$E(Y \mid Y^* + \sum_i \varepsilon_i h_i(A_i \mid W_i))$$

Here *Y* represents walking speed, and Y^* represents the predicted walking speed based on the *FX* model described previously; $\sum_i \varepsilon_i$ represents the sum of the estimated difference in *Y* as the result of confounding of Lnfat and LTPA by other variables, not fully captured by Y^*. Based on the estimated model above, mean walking speed was predicted for different combinations of LTPA(=0,1), representative of population-level participation in < 22.5 Mets/week and $\geq$ 22.5Mets/week, respectively; and Lnfat(=0,1) , representative of

population-level measures of lean mass to fat mass $\leq$ sex-specific median and > sex-median, respectively---e.g.,

$$E\left(\hat{Y}(A_1 = 0, A_2 = 0)\right) = \frac{1}{n}\left(\hat{Y}^* + \hat{\varepsilon}_1 h_1(0, W_1) + \hat{\varepsilon}_2 h_2(0, W_2)\right)$$

$$\text{where } \hat{Y}^* = \sum \hat{\beta}_0 + \hat{\beta}_1 I(A_1 = 1) + \hat{\beta}_2 I(A_2 = 1) + \hat{\beta}_p * Co\,\text{var}\,iates_p$$

Influence curve-derived 95% Confidence Limits of TMLE estimates. For mean estimates of walking speed, variance estimates were computed based on influence-curve calculations as described by Gruber and van der Laan (1) . For the mean walking speed estimate above, the influence curve estimate would be

$$IC = \left(Y - \hat{Y}^*\right) * h_1(0, W_1) * h_2(0, W_2) + \left(\hat{Y}^* + \hat{\varepsilon}_1 h_1(0, W_1) + \hat{\varepsilon}_2 h_2(0, W_2)\right) - E\left(\hat{Y}(A_1 = 0, A_2 = 0)\right)$$

$$\text{and } \hat{\sigma}^2 = \frac{1}{n}\sum_n IC^2$$

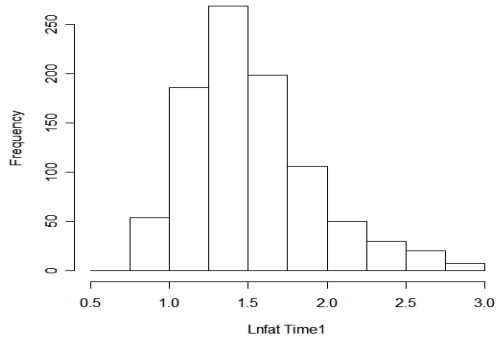
$$\text{and estimated SE for } E\left(\hat{Y}(A_1 = 0, A_2 = 0)\right) = \hat{\sigma} / \sqrt{n}$$

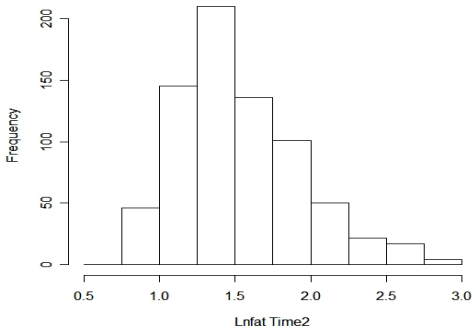
Further details with regard to TMLE are provided by van der Laan and Rubin elsewhere (2).

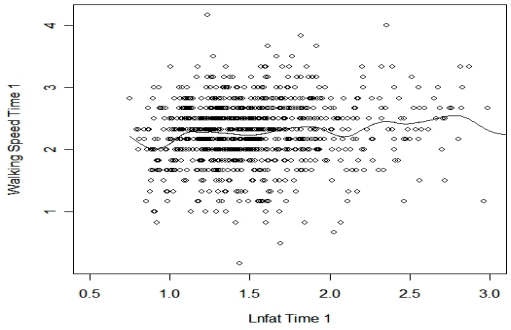
References

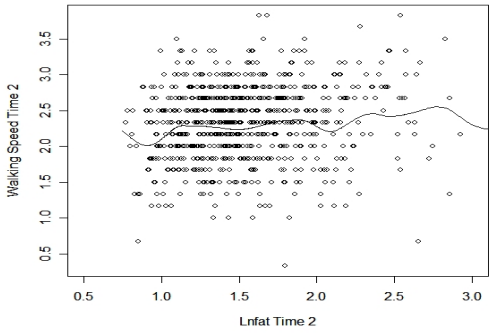
1. Gruber S, van der Laan M. Targeted Maximum Likelihood Estimation: A Gentle Introduction. Berkeley Electronic Press 2009;252.
2. van der Laan M, Rubin D. Targeted maximum likelihood learning. International Journal of Biostatistics 2006;2.

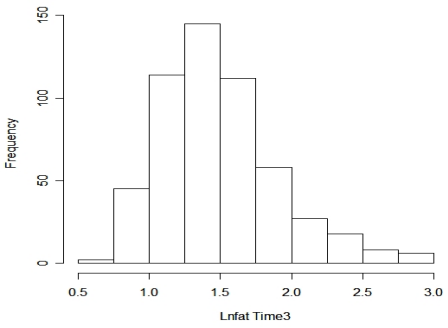
Histograms of Lnfat and scatterplots of Walking Speed vs. Lnfat with smooths in females, time 1-4.

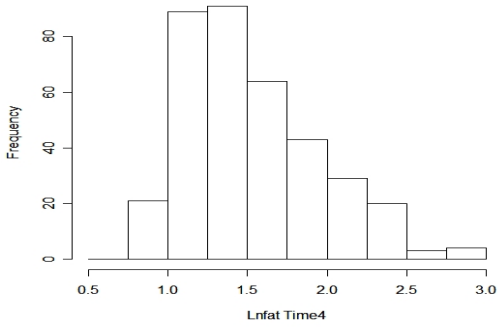


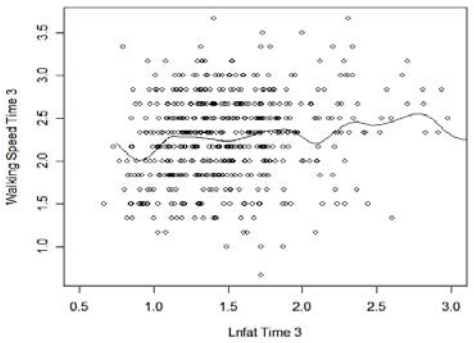


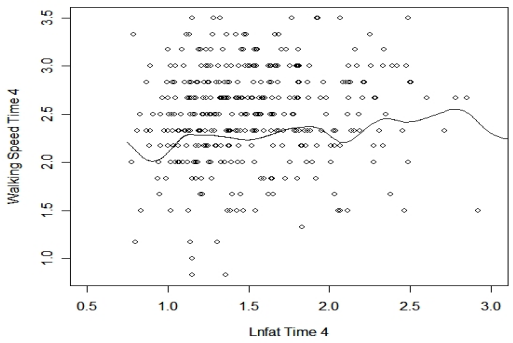












Histograms of Lnfat and scatterplots of Walking Speed vs. Lnfat with smooths in males, time 1-4.