Sustainable Intensification Assessment Methods Manual (Working Draft)

24 October 2017

By

Mark Musumba (UF[1]), Philip Grabowski (MSU[2]), Cheryl Palm (UF[1]), and Sieglinde Snapp (MSU[2])

Steering Committee:

Gundula Fischer (IITA[3]), Bruno Gerard (CIMMYT[4]), Jerry Glover (USAID[5]), Fred Kizito (CIAT[6]), Generose Nziguheba (IITA[3]), Vara Prasad (KSU[7]), Peter Thorne (ILRI[8]), and Bernard Vanlauwe (IITA[3]).

Acknowledgments:

We are grateful to the Feed the Future Innovation Lab for Sustainable Intensification, Africa Research in Sustainable Intensification for the Next Generation (AfricaRISING), and Consortium for Improving Agriculture-based Livelihoods in Central Africa (CIALCA) researchers for their support and contributions during the field visits and trainings.

This study is made possible by the support of the American People provided to the Feed the Future Innovation Lab for Sustainable Intensification (SIIL) through the United States Agency for International Development (USAID). The contents are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government. Program activities are funded by the United States Agency for International Development (USAID) under Cooperative Agreement No. AID-OAA-L-14-00006.

Affiliations:

[1]University of Florida; [2] Michigan State University; [3]International Institute for Tropical Agriculture; [4]International Maize and Wheat Improvement Center; [5]United States Agency for International Development; [6]International Center for Tropical Agriculture; [7]Kansas State University; [8]International Livestock Research Institute.

Table of Contents

SI Indicator Manual

The sustainable intensification indicator manual provides guidance on how scientists working in research for development projects can estimate the indicators and metrics listed in the Guide for Sustainable Intensification Assessment. This manual provides the most commonly used methods to estimate an indicator or metric and, where possible, one or more alternative approaches.

For each domain, we provide a summary table of the indicators, metrics, and methods of measurement organized at four different spatial scales. The tables are followed by detailed explanations of the indicators, metrics, and estimation procedures. We define an indicator as a “quantitative or qualitative factor or variable that provides a simple and reliable basis for assessing achievement, change or performance” (ISPC, 2014). Metrics are computed by aggregating and combining raw data (e.g., yield or height-for-age). They often represent the values on which indicators are built.

For each metric, we provide one or more methods of data collection. For each of these methods we describe the following information:

- the data needed to estimate the indicator and the method of data collection,
- the unit of analysis and algorithm that can be used to compute the indicator, and
- the limitations in estimation and interpretation of the method or metric.

At the end of each indicator section we provide a summary of the limitations to estimation and interpretation that are common to all metrics for that indicator.

Productivity Domain (Note: The superscript letters (a,b,c) after each metric refer to the measurement methods in the right-hand column)

PRODUCTIVITY DOMAIN					
Indicator	**Field/plot level metrics**	**Farm level metrics**	**Household (hh) level metrics**	**Community/ Landscape + metrics**	**Measurement method**
Crop productivity	Yield (kg/ha/season) [a,b,c](including tree product/area under crown) Rating of yield [d]	Yield (kg/ha/season) [a,b,c]		Remotely sensed measures of crop productivity (kg/ha/yr) [e]	[a] Yield measurements [b] Recall survey [c] Crop models [d] Farmer evaluation [e] Remote sensing
Crop biomass productivity	Forage production (kg/ha/season) [a,b,c] Residue production (kg/ha/season) [a,b,c] Rating of residue production [d]	Residue production (kg/ha/season) [a,b,c]		Remotely sensed measures of crop biomass (kg /ha/yr) [e]	Same as for yield
Animal productivity	Animal products (amt./animal/yr.) [a,b] Animal by-products (amt./animal/yr.) [a,b] Rating of animal productivity [c]	Animal product per unit land (amt./ha/yr) [a,b] Animal byproduct per unit land (amt. /ha /yr) [a,b] Herd composition [a]	Animal product per hh (product/hh/yr) [a,b] Animal byproduct per hh (product /hh/yr) [a,b]	Net commercial offtake (product /ha/yr) [a]	[a] Recall survey [b] Production measurements [c] Farmer evaluation
Variability of production	Coefficient of variability [a] Probability of low productivity [a]	Coefficient of variability [a] Probability of low productivity [a]	Rating of variability [b] Rating of production risk [b]	Variability of NPP [c]	[a] Productivity over time [b] Farmer evaluation [c] Remote sensing
Input use efficiency	Product per input [a,b]				[a] Survey and productivity measures [b] Models
Yield gap	Yield gap per crop (kg/ha/season)	Yield gap per crop (kg/ha/season)		Yield gap (kg/ha/season)	Same as Yield
Cropping intensity	# of cropping seasons per year [a]			#of cropping seasons per year	[a] Recall survey
Post-harvest losses			% harvest lost [a,b]		[a] Direct measurements [b] Survey

Productivity Domain

Crop Productivity

Description of the indicator:

Crop productivity is a measure of the total sum of annual plant production, which is also known as net primary productivity. Crop productivity can be partitioned by tissue type (e.g., grain, leaves, and stems) based on how the plant is used. The unused portions of crops are often referred to as crop residues, which is the next indicator.

Metric 1: Yield

Description of the metric

Yield is a measure of crop production for a given land area, generally measured at the field scale. Yield typically focuses on a limited portion of the plant, such as the grain for row crops. However, in many cases farmers use nearly all parts of the plant for various purposes. For this reason, we suggest taking into consideration all plant parts used by farmers, remembering that stover left in the field is often consumed by livestock. In many cases, it may be reasonable to focus on grain yield and stover. The portion left in the field and not consumed by livestock plays an important role in nutrient cycling and this is measured in the Crop Residue Productivity Indicator.

Measurement method 1: Yield cuts

Method of data collection and data needed

Yield cuts are when a destructive harvest is carried out to measure production at crop reproductive maturity for a known area of land. For indeterminate reproductive plants this will require multiple harvests. For tree and bush species, yield can be estimated by using algorithms of stem (trunk) diameter at a specific height related to overall tree product per area under crown. The biomass produced can be partitioned by tissue type, such as grain and stover.

Yield cuts are common in agronomic trials and can be used in combination with household surveys. They are typically performed by randomly selecting a location from an experimental plot or from a plot on a farm and cutting the plants from a measured area (e.g., one square meter). In some cases, the entire harvest of a field with a known area will be measured. For guidelines on sampling a plot with quadrats and rating the vegetation within those plots, see Anderson and Ingram (1993, section 3.1.2, which is copied in the Crop Residue Productivity Indicator under measurement method 1).

The biomass collected for grains are typically dried to standard moisture (e.g., 15.5% for maize in the U.S.) before being weighed. This can be accomplished by drying the grain in the sun or in an oven at low temperatures. Alternatively, the weight can be adjusted if the moisture content is known.

Unit of analysis

The unit of analysis is the dry weight of plant biomass, grain, and stover per area of land.

A measured value of 100 grams of grain per square meter is equivalent to 1 metric ton per hectare.

Limitations regarding estimation and interpretation

Plant populations often vary across a single plot and this needs to be considered if sampling a portion of the plot for a yield cut. Plant population density is especially important when estimating yield by extrapolating the measured production of a few plants. Measuring plant population density is preferably done by counting the plants in a known area at harvest time. In some conditions, plant population density can be estimated by the seeding rate (seeds planted per hectare), but this assumes high germination rates and low plant mortality. Knowing the plant population density is also important for being able to interpret the reasons for yield changes or differences.

Measurement method 2: Farmer recall of yield

Method of data collection and data needed

Household surveys or field-based farmer surveys are opportunities where farmers are asked to report their production for a given plot and the area where the crop was produced. During these agricultural surveys, total production from the farm for each crop grown may also be ascertained. The questions are mainly enumerated after harvest or at the end of the agricultural year. The section on the area planted is measured and for each field the area harvested is also collected per crop.

Data on amount harvested are collected and the units (bags, ox-carts, baskets) are recorded by the enumerator. It is necessary to ensure that a conversion factor for the units is obtained (e.g., five 90-kilogram bags of maize). In this case, the weight can be converted to a common unit like kilograms. To prevent under-estimation of yield, enumerators should also ensure that for perennial crops or some annual crops, they ask about the amount of unharvested planted crop still in the field .

Unit of analysis:

The unit of analysis is the farmer's estimated weight per area, or volume per area.

Limitations regarding estimation and interpretation

Farmer recall of the amount harvested may be reasonably reliable but are typically in local units (ox-carts, baskets, sacks), which then need to be accurately converted to kilograms.

In most contexts, farmer estimates of area are not very accurate (Fremont and Benson, 2011). Farmers may overstate their land holding because land is a source of prestige. In other cases, farmers may under report land owned for fear of being targeted for land redistribution. Measuring field size is important and can now easily be achieved by walking the boundary with a handheld GPS unit which can precisely calculate the area in hectares.

Measurement method 3: Crop modeling

Method of data collection and data needed

Crop models can be used to estimate yields for a wider range of environments, including edaphic (soil) or weather conditions not observed during the experiment.

Various crop models are commonly used for this purpose. Two of the most common are open-source software with strong support for global use, namely DSSAT (Decision Support System for Agrotechnology Transfer, http://dssat.net/downloads/dssat-v46) and APSIM (Agricultural Production Systems sIMulator, https://www.apsim.info/AboutUs.aspx). Peer reviewed publications are available describing these models (for DSSAT, see Jones et al., 2003; for APSIM see Holzworth et al., 2014).

Crop models simulate crop growth, development, and yield of a crop growing on a uniform area of land under prescribed or simulated management, as well as the changes in soil water, carbon, and nitrogen over time. The minimum data needed to run these models typically include:

1) daily weather data (max temp, min temp, precipitation, and, if possible, solar radiation);
2) soil characteristics (soil texture, N, C, cation exchange capacity, pH, at several depths up to 50cm);
3) crop phenology by variety/cultivar in a given environment (the timing of emergence, canopy development, anthesis [flowering], and maturity); and
4) management practices (planting date, plant spacing, row spacing, fertilization dates), fertilizer types and amounts, harvest date, weeding and ploughing dates, ploughing and weeding implements, ploughing depth, harvested grain weight (dry), harvested biomass weight (dry).

Daily weather data can be obtained from nearby meteorological stations, many of which are shared publicly among crop modelers. Similarly, data on soil characteristics are available from the World Soil Information System (http://www.isric.org/projects/data-wosis-project) which includes the ISRIC-WISE global soil profile database (https://daac.ornl.gov/SOILS/guides/Isric.html). Crop models may include botanical parameters for well-studied varieties but further data may need to be collected for under-studied environments and varieties. The data for observed botanical traits and the data on common management practices can be collected through simple surveys for use at the household level (e.g., Mourice et al., 2014).

Unit of analysis

The unit of analysis is the plant biomass, grain, and stover dry weight per area of land. The accumulated nutrients, or carbon in plant biomass, can also be quantified by this method.

Limitations regarding estimation and interpretation

Crop modeling requires considerable investment in data collection for calibration and validation of local conditions.

Metric 2: Farmer rating of yield

Description of the metric

In cases where farmers have tried a new technology but on-farm yield measurements and detailed surveys are not available, it is possible to obtain farmers' qualitative evaluations of yields, for example asking them to rank or rate yields for varieties or management practices.

Method of data collection and data needed

Farmer rating of yield can be operationalized in many ways. For example, farmers can be asked how they rate the yield of a given management practice or variety on a scale of 1 (very poor) to 5 (very good). Figure 1 provides an example with four categories (0-3) which are then further subdivided into four levels (a-d). This particular example is designed for use with cards representing various intercropping systems and species combinations, but it could be adapted even for varieties or other management practices.

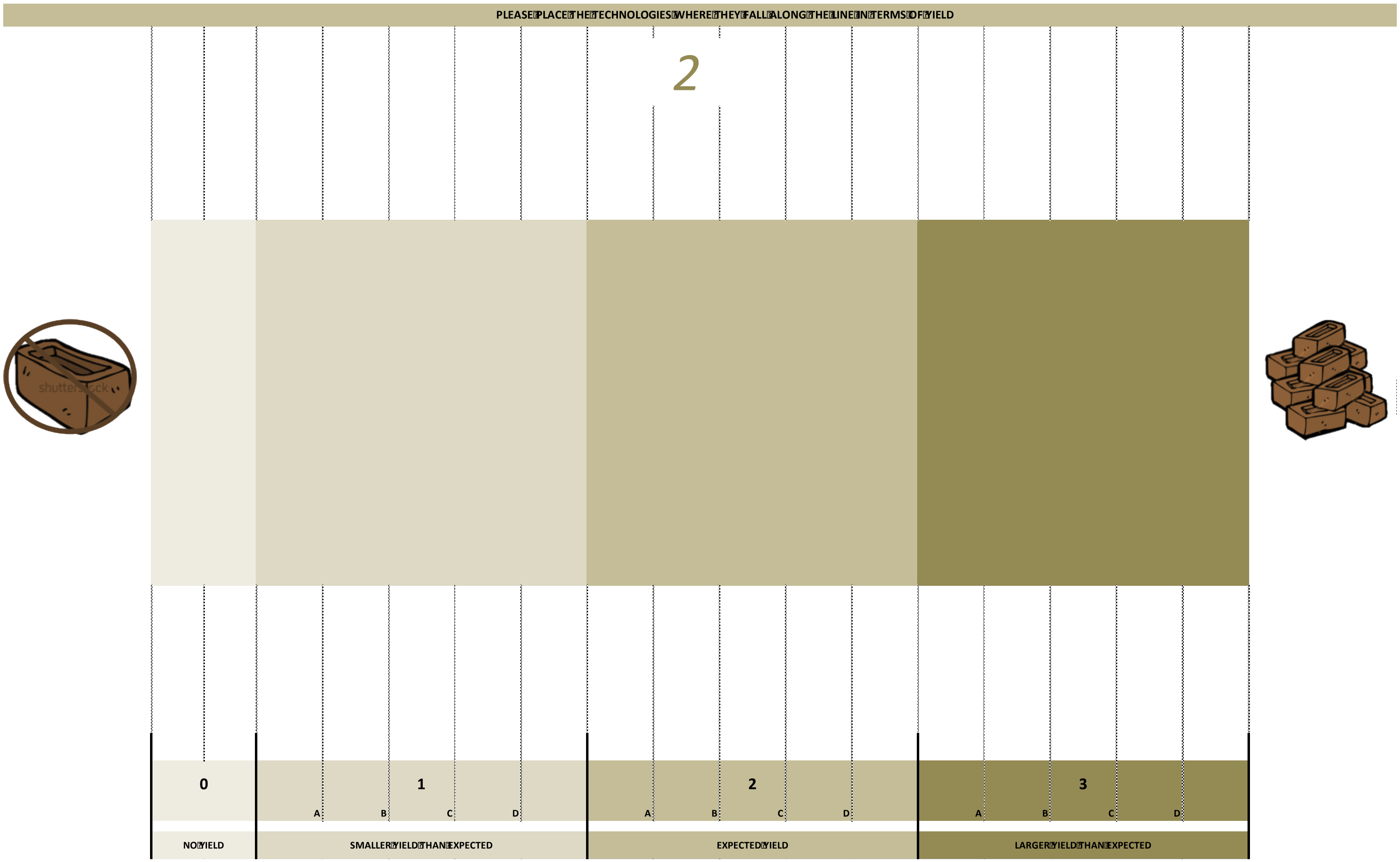
The amount produced and the area on which it is produced are not usually made explicit in this type of rating. Farmer ratings may be based on implicitly comparing new technologies with what they believe would have been produced from their own practice on that field during the season the new technology was used.

A first step in analyzing these data is to create a table comparing the frequencies of each score for all comparisons (e.g., technologies, varieties) or graphing those frequencies (Coe, 2002).

Coe (2002) cautions against treating these scores as if they were measured on a continuous scale (i.e., comparing means, carrying out ANOVAs or using them as continuous variables in a linear regression).

Instead, he suggests using a statistical program to model the information as ordered categorical data, which allow for ordinal regressions. More details about analyzing rating data (which is ordered categorical data) can be found in Coe (2002), and an empirical example is found in De Groote et al., 2010.

Figure 1. Card for farmer rating of yield (Source: Erin Anders, 2016. Participatory Action Research Tools: Likert Scale Rating with Small Holder Farmers)



Unit of analysis

The unit of analysis is the individual responding to the questions.

Limitations regarding estimation and interpretation:

Rating of yield is a coarse estimate of actual production, which may be more subjective than physical measurements. Where possible, ratings should be disaggregated by, for example, gender or wealth categories, to see what differences exist from these diverse perspectives.

Metric 3: Remotely-sensed measures of crop productivity

Description of the metric

Remote sensing can be used to measure vegetation indices and productivity across landscapes through spectral signals and multi-input algorithms. One well known index is the Normalized Difference Vegetation Index (NDVI), which has the following formula: NDVI = (NIR−R) / (NIR+R), where NIR is Near Infra-Red and R is Red. NDVI can be used to measure crop greenness, i.e., vegetation density and health, across both spatial and temporal scales. Another, Net Primary Productivity (NPP), is a multi-input modeled product that estimates kilograms of carbon per square meter. Measures of carbon per unit area, in combination with harvest indices, can be used for direct conversion to plant biomass (i.e., yield). Another more complex method is to use crop models to estimate yield given the remotely-sensed leaf area index (LAI) estimates over the growing season (Lobell, 2013).

Method of data collection and data needed

The two most commonly used sources of remotely sensed agricultural data are from U.S. government funded satellites—LANDSAT (managed by USGS) and MODIS (managed by NASA).

LANDSAT provides fine resolution images (30-m by 30-m pixels) at 8-day intervals from 1972 and are released about 24 hours after acquisition (USGS, no date) from the following website: https://landsat.usgs.gov/landsat-level-1-standard-data-products. This resolution makes it possible to detect fields roughly 1 ha or larger (Lobell, 2013). To calculate NDVI, it would be necessary to download the near infrared (band 4 of Landsat 7) and red wavelengths (band 3 of Landsat 7).

MODIS (Moderate Resolution Imaging Spectroradiometer) provides coarse resolution images (250-m to 1-km) of the earth's surface every 1 to 2 days (commonly composited into 8- or 16-day intervals), beginning in 1999. Reeves et al. (2005) used MODIS Gross Primary Productivity (GPP) data to estimate wheat yields in the U.S. and found it accurate for estimating state-level production. Peng et al. (2014) used MODIS NPP data to estimate paddy rice yields in China by complementing the remote sensing with field visits. Messina et al. (2017) used Net Photosynthesis (PsnNet), a derivative of (GPP), to evaluate crop productivity in the Malawian smallholder farming context.

The data from these satellites can be downloaded as raster structure data, which can then be manipulated using free GIS software, such as QGIS, GRASS GIS, or the statistical software R. A useful R tutorial on raster structure data can be found here (https://geoscripting-wur.github.io/IntroToRaster/).

Unit of analysis

The unit of analysis is the area of interest, with a lower limit being set by the resolution of the image, which determines the dimensions of each pixel detected.

Limitations regarding estimation and interpretation

Inferring from remote sensing to estimate grain production from agricultural landscapes can be difficult, especially where small plot sizes create a diverse mosaic of crops and management practices within the smallest spatial resolution of the satellite image. Many sources of remote sensing have images with a resolution around 30 to 100meters, whereas yield variation is generally at a 1- to 10-meter resolution.

This limitation can be overcome using commercial satellites, planes, or drones. Lobell (2013) states, "new commercial systems are delivering even higher spatial resolution (5m × 5m or finer) at costs that are approaching 1 United States dollar(USD) per km^2 (or \$0.01 per ha)".

Shortcomings regarding estimation and interpretation of the indicator

The inference zone is important to consider for all crop productivity metrics. On-farm measurements, and farmer qualitative data, may be influenced by local conditions and culture, thus extrapolation from this information needs to be carefully considered.

The quality of production is often important and this can be difficult to account for in a quantitative measure of kilograms per hectare. One approach is to compare production for similar quality of product from the same species (for example, dividing potato production into market grades). Fodder quality assessment is described in the Crop Residue Productivity Indicator section.

Relative differences in productivity are even more complex to measure when multiple crops are produced on the same field. In some cases, the production of all crops can be compared in terms of the monetary value of production (which relates to income, an indicator in the economic domain) or their nutritional value (e.g., calories, protein), which is an indicator in the human condition domain. More often it is necessary to report production separately for each of the different types of plants. Productivity of intercropping is typically compared to monocropping through a calculation known as the land equivalent ratio (Mead and Willey, 1980):

$$LER = \frac{Y_a}{S_a} + \frac{Y_b}{S_b}$$

where Y_a is the intercropped yield of crop A, S_a is the sole crop yield of crop A, Y_b is the intercropped yield of crop B, and S_b is the sole crop yield of crop B. A LER of 1.2 means that it would take 1.2 hectares of sole crops to produce the same amount as 1 hectare of the intercrops.

Comparisons of productivity of the land at the farm or landscape level requires aggregating across crops (e.g., cotton vs. maize). This aggregation must convert the production of each crop into a common unit that is meaningful for farmer decision-making (such as local currency, calories, or protein). Alternatively, farmers' subjective valuation of each crop could be used to assign weights to each species. This may be particularly useful if a large portion of the crops is not sold.

Further reading on additional methods for estimating crop productivity are available from the CCAFS SAMPLES website (http://samples.ccafs.cgiar.org/measurement-methods/chapter-8-yield-estimation-of-food-and-non-food-crops-in-smallholder-production-systems/) which is also available as a book chapter (Sapkota et al., 2016).

Crop Biomass Productivity

Description of the indicator

For sustainable intensification, the productivity of the land needs to be assessed in terms of all that is produced (not just grain yield). This is especially important where residues are used for fodder or returned to the soil.

Metric 1: Residue production

Description of the metric

Crop residue productivity is a measure of the 'non-grain' biomass from plant production on a known area, which is also the difference between Net Primary Productivity and yield.

Measurement method 1: Biomass measurements

Method of data collection and data needed

Crop residues can be weighed at harvest from a known area using the methods described in the Tropical Soil Biology and Fertility manual (Anderson and Ingram, 1993) Sections 3.1.2 to 3.2.2 which are copied in Box 1.

Box 1. TSBF section 3.1.2 Herbaceous plants and short duration crops
Total biomass is measured by harvesting, drying, and weighing a number of small subsamples, or quadrats. Quadrat size depends on plant spacing: 0.5 m x 0.5 m is a convenient size for most grasslands; 1 m x 1 m may be appropriate for crops such as maize. Sample number (n) should be sufficient to reduce the standard error to about 10% of the mean. Use of 20 to 30 samples per treatment, distributed amongst replicates, is usually sufficient. Sample location should best be random, but systematic with a random start is acceptable.
Procedure
Cut all herbaceous vegetation within the quadrat at 2 cm above the ground (to avoid soil contamination), and sort into live (green) biomass and standing dead if possible.
Collect the litter from the ground for an estimate of litter standing crop.
Dry all samples as soon as possible to prevent decomposition.
Species composition in mixed communities is estimated by the dry-weight-ranking technique. The technique is based on a multiple regression for the dry weight of a mixed sample of herbage on the weights of the three heaviest species in the mixture. Experience indicates that it is easier to assess visually the rank order of the species in a quadrat than to estimate accurately their biomass. Tests in a large number of different communities have shown that the regression coefficients are fairly consistent between communities, and therefore do not need to be recalculated each time. The technique does not work well in communities completely dominated by one species, and tends to ignore rare species. The modified form given here, where the total biomass within the quadrat is also given a visual score, gives better results in communities where the total biomass is patchily distributed. Quadrat size should be small enough that species ranking is simple, but large enough that most quadrats have at least three species in them. Quadrats 0.5 m x 0.5 m square are usually adequate in grasslands. About 50 quadrats should be assessed per treatment.
Procedure
Walk around the plot to obtain a clear visual impression of what the minimum (1) and maximum (5) quadrat biomass looks like.
Locate the quadrats randomly or systematically after a random start.
In each quadrat (i) give the total biomass a score (w) between 1 and 5 according to whether it is near the minimum or maximum for the plot.
In each quadrat, give the species (j) which contributes most to the total quadrat biomass a rank score (r_{ij}) of 1, the second heaviest species a rank of 2, and the third heaviest species a score of 3. If a single species contributes more than about 70% of the biomass, give it ranks 1 and 3, or 1 and 2 (or even 1, 2 and 3 if it is the only species in the quadrat). Similarly, the second species could get a 2 and 3 if necessary.
(Excerpted from Anderson and Ingram, 1993, pp. 27-28)

When all the quadrats have been scored and ranked, calculate the score for each species:
Calculation

$$Species\ score = 70.2\ {}_r\sum w_i +\ 21.1\ {}_r\sum w_i +\ 8.7\ {}_r\sum w_i$$

where Σw_i is the sum of the quadrat scores for the quadrats where species j obtained rank r.
Add up the species scores to give a total score.
Determine the % contribution by species j to the total biomass:
Species j contribution to biomass (%) = (species score / total score) x 100.
Further reading

Gillen, R.L. and Smith, E.L. (1986) Evaluation of the dry-weight-rank method for determining species composition in a tallgrass prairie. Journal of Range Management 39, 283-285.
Jones, R.M. and Hargreaves, J.N.G. (1979) Improvements to the dry-weight-rank method for measuring botanical composition. Grass and Forage Science 34, 181-184.
Sandland, R.L., Alexander, J.C. and Haydock, K.P. (1982) A statistical assessment of the dry-weight-rank method of pasture sampling. Grass and Forage Science 37, 263-272.

3.2 ABOVE-GROUND INPUTS
3.2.1 Tree and shrub litter
In a comprehensive review of tropical litter fall data, Proctor (1983) observed that the results of many published studies were not comparable. This resulted from inadequate siting and replication of traps in relation to site heterogeneity, sampling for periods of less than a year and lack of standardization of small litter fractions (laves, twigs, reproductive structures and “trash”).
Litter trap construction
Litter traps are bags or boxes supported just clear of the ground with an aperture of 0.25- 1 m^2. A circular construction is best as it minimizes edge effects. Woven plastic bags are light weight for use in remote sites and can be tensioned into shape using lines attached to D-rings sewn around the mouth of the bag. The traps must allow free drainage of rain water but have a mesh size of approximately 1 mm or less to retain fine litter fractions. Trays on the ground surface can be used to measure litter-fall from dwarf shrubs etc., but animal activity, drainage and wind can present problems.
Similar considerations apply to collections of litter from quadrats on the ground, which may be necessary for estimating falls of palm fronds and larger woody litter. Trash fractions, however, which often have low mass but high nutrient content, will be lost by this method.
Procedure
Randomly locate litter traps (for material other than branches) within moderately homogeneous plots, or in a stratified random pattern (with 10 traps per subplot) in sites where it is necessary to include major variation in topography, soils and vegetation structure. (Excerpted from Anderson and Ingram, 1993, pp. 29 - 30)

Note: To achieve a 5% standard error about the mean, Newbould (1967) recommends the use of at least 20 traps/plot. In very heterogeneous sites, however, higher numbers of traps may be required.
Collect litter every 2 weeks and air dry it. More frequent collections may be necessary for litters which decompose rapidly, e.g. some tree legumes, while less frequent collections may be made under dry conditions (though the possibility of the litter becoming contaminated with dust and/or animal faeces should be recognised). [If the information in this box is taken verbatim from another document, it should be cited]
Sort the dried material into:

- leaves (including petioles and foliar rachises);
- small woody litter (twigs < 2 cm in diameter and bark);
- reproductive structures (flowers and fruits could be differentiated);
- trash (sieve fraction < 5 mm).

(For palm fronds, the leaflets, the rachis below 2 cm, and the remaining parts of the rachis should be weighed and recorded separately.)
Oven-dry subsamples of litter to obtain correction factors for moisture content (see Section 6.1)
Express all fractions defined above on an oven dry basis in g/m2/year or t/ha/year with 95% confidence limits.
Estimate branch fall from large (e.g. 100 m2) ground quadrats. Break twigs at the 2 cm diameter point, weigh the > 2 cm diameter material, subsample for oven-dry mass and other determinations as required.
3.2.2 Herbaceous litter and above-ground crop residues
The minimum level of sampling is at maximum and minimum biomass associated with major seasonal changes or perturbations; i.e. sampling four times a year under climatic regimes with a strongly bimodal pattern of rainfall. This will underestimate litter inputs as a consequence of material turning over between sampling dates and sampling at regular intervals every few weeks is recommended.
Procedure
Determine herbaceous litter (including grasses and forest ground flora) by harvesting quadrats in conjunction with biomass estimates (Section 3.1.2).
Separate litter, where possible, by plant species for the most frequent 80 % of species and bulked for the remaining 20%. (This may be impractical in very species-rich communities.)
Determine crop residues after harvest and at the time of ploughing or other manipulation.
Oven-dry litter subsamples to obtain correction factors for moisture content (see Section 6.1). Litter heavily contaminated by soil may need to be corrected for 'ash' content as well.
Express all fractions defined above on an oven dry basis in g/m2/year or t/ha/year with 95% confidence limits.
(Excerpted from Anderson and Ingram, 1993, pp. 29-30)

Unit of analysis

The unit of analysis is the dry weight of plant biomass per area of land. A measured value of 100 grams per square meter is equivalent to 1 metric ton per hectare.

Limitations regarding estimation and interpretation

One challenge with directly measuring the total amount of residue biomass is that if measurement is done once at harvest then it may miss the biomass from plants that lose leaves through the growing season, which may rot before harvest. Leaf traps collected monthly or weekly can be used to measure such biomass (see TSBF section 3.2.1 copied above). This may be especially important for establishing linkages between productivity and soil organic matter.

Measurement method 2: Farmer recall of residue production

Method of data collection and data needed

In some contexts, farmers may be able to estimate the amounts of crop residues produced, such as when the residues are cut and stored for livestock feed. In such cases, crop residues can be estimated by farmers for a measured land area in a similar way as yield estimates are carried out (see "Crop productivity" indicator for more details).

Unit of analysis

The unit of analysis is the dry weight of plant biomass per area of land. A measured value of 100 grams per square meter is equivalent to 1 metric ton per hectare. Farmers are likely to share residue production using local units, such as ox-carts, bales or heaps. These can be approximately converted to kg by carefully measuring the weight of several local units.

Limitations regarding estimation and interpretation

In most contexts, farmers are not able to quantitatively estimate crop residue production in any reliable way, especially where residues are left in the field. In such contexts, it is better to consider another method or metric (such as farmer rating of residue production).

Measurement method 3: Using grain yield as a proxy for biomass

Method of data collection and data needed

If the residue biomass is not measured but grain yield is measured, then the residue biomass can be estimated from the harvest index for the variety of the crop. The harvest index is simply the portion of all biomass that is harvested as grain. As with yield, the crop residue biomass can be estimated through remote sensing of NPP as the non-grain portion of NPP, usually inferred by the harvest index.

Unit of analysis

The residue biomass can be estimated as follows:

$$R = (G - G*HI)/HI$$

where R is the residues, G is the grain harvested, and HI is the harvest index. For more details on harvest index see Kawano (1990).

Limitations regarding estimation and interpretation

The harvest index is not a fixed attribute of a crop but can vary across environmental conditions. For this reason, using grain yield to estimate biomass is a rough approximation, not a precise calculation.

Metric 2: Rating of residue production

Description of the metric

Farmer rating of crop residue production may be possible even in contexts where farmers are unable to estimate the quantities produced. Farmers' ratings are more likely to have accurate information where residues are actively used, such as for fodder or fuel. The methods for carrying out and analyzing the ratings are described in the Crop Residue Productivity Indicator section.

Unit of analysis

The unit of analysis is the individual responding to the questions.

Limitations regarding estimation and interpretation

Rating of residue production is a coarse estimate of actual production, which may consist of more subjective than physical measurements. Where possible ratings should be disaggregated by gender, wealth categories, etc., to see what differences exist from these diverse perspectives.

Metric 3: Remotely-sensed measures of crop Biomass

Description of the metric

As described in Metric 3 of the Crop Productivity Indicator, remote sensing can be used to estimate crop growth. See that section for further details on how to estimate production with remotely-sensed data.

Unit of analysis

The unit of analysis is the area of interest, with a lower limit being set by the resolution of the image, which determines the dimensions of each pixel detected.

Limitations regarding estimation and interpretation

Inferring from remote sensing to estimate residue production from agricultural landscapes can be difficult, especially where small plot sizes create a diverse mosaic of crops and management practices within the smallest spatial resolution of the satellite image. Many sources of remote sensing have images with a resolution around 30 to 100m, whereas crop production variation is generally at a 1- to 10-m resolution.

This limitation can be overcome using commercial satellites, planes, or drones. Lobell (2013) states, "new commercial systems are delivering even higher spatial resolution (5-m × 5-m or finer) at costs that are approaching 1 USD per km^2 (or $0.01 per ha)".

Shortcomings regarding estimation and interpretation of the indicator

For all crop residue metrics, the quality of crop residues is important to consider. When its primary use is as animal feed, quality should be directly assessed for attributes important to the specific animal system (e.g., digestibility, crude protein, nutrients). The simplest way to assess how a change in species would affect residue quality for animal feed is to ascertain the average nutritional quality of each fodder (using the most prevalent species if mixed). A more robust method is to use Near Infrared Spectroscopy (NIRS), which is more affordable than wet chemical testing (Stuth et al., 2003). This requires a clean vegetation sample (no soil) and it must be dried soon after sampling. A detailed manual (though dated) for this method is available online (Marten et al., 1989).

If residues are incorporated into the soil, it may be useful to analyze the amount of nitrogen and other plant nutrients to determine potential impacts on soil fertility. The ratio of carbon to nitrogen in the residues is an important factor in the rate of decomposition, which determines when the nutrients are likely to be available for the plants. The C:N ratio can be estimated through lookup tables. Nutrient values of tropical species have been developed through the Organic Resource Database (Palm et al., 2001).

Animal Productivity

Description of the indicator

Animal productivity is the total sum of products and services derived from animals. In the context of sustainable intensification, the efficiency of that production is important, for example, the amount of land required to produce the feed for the animals. Though there are dozens of species of animals raised for various purposes, the great majority of products are derived from the primary livestock species – cattle, pig, sheep, goats, and poultry. Analyses focusing on other species are certainly possible and may be essential in some contexts.

Quantifying animal productivity can be complicated because of the multiple uses for animals. To understand the multiple uses for each type of animal, Dorward et al. (2005) provide a simple method for creating a matrix with animals as columns and functions as rows. These functions include consumption and sale of typical animal products, as well as more difficult to quantify functions, such as savings, insurance, and buffering. Moss et al. (2016) recommend tailoring the analysis of production to the objectives of the producers by categorizing livestock producers into classes, such as subsistence, commercial, and patio producers.

Another challenge in measuring productivity is that animal mobility can make it difficult to quantify their feed sources. Metric 1: Animal Products and By-Products focuses on estimating total production or production per animal. Production density (product per land area) can be calculated as total production (kg/year) divided by the product of feed productivity (kg/ha) and feed consumption (kg/year). A complete analysis will consider all the products from the animals (including manure, draft power, and transportation services).

Metric 1: Animal products and by-products

Description of the metric

Animal productivity can be assessed as production per animal (e.g., eggs per chicken per month), production density (quantity per unit farm land), or total production (e.g., kg of meat). Production per animal is listed at the “plot” scale. The farm scale is where production density is listed (production per land area). The total production per household is listed at the household scale. To estimate community or landscape scale production (or wider), sampling is an important issue. Moss et al. (2016) note that livestock may not be distributed evenly with households, especially in countries where livestock are concentrated in areas with low human population density, such as semi-arid regions. Transhumant and nomadic herders can be difficult to reach for surveys. Timing of data collection is also important to consider because animal production and herd size may vary seasonally, and cultural events may cause high demand in a short period of time (Moss et al., 2016).

Table 2 presents a list of common animal products and guidelines about collecting data on them.

Table 2.. Guidelines for measuring animal productions (adapted from Moss et al., 2016).

Animals	Product	Calculation	Data collection guidelines
All	Various	# of animals slaughtered	Divided by type (i.e., lambs, ewes, rams or calves, cows, bulls) and by gender (except for poultry)
Cattle, pigs, goats, sheep	Meat	# of animals slaughtered * avg. weight	Annual survey may be adequate
	Hides	# of animals slaughtered * # of skins per slaughtered animal	Salted and unsalted hides
Cattle, goats	Milk	Mean # lactating animals * avg. daily milk production * 30 days * avg. # months each animal produces milk	Herd size can be used to estimate production if the following variables are known: proportion of cows, percentage of cows that lactate per year, average output per lactation, and level of extraction for human use (not suckled by calves) Proxy indicators for dairy productivity include age of cow at first calving and the interval between calving
Poultry	Meat	# of animals slaughtered * avg. weight	Frequent data collection is needed to accurately estimate poultry inventory and animals slaughtered
	Eggs	Average # eggs per bird * avg. # birds laying	Average eggs per bird should be for a given period, such as a week or month
Sheep	Wool	# of animals * avg. weight per animal	
Cattle, donkeys, horses, camels, etc.	Draft	Total value, days, or area of draft per draft animal	Assessment of the condition of the draft animals before, during, and after their peak season provides further information about efficiency
	Transport	Total value from transportation or weight * distance transported per draft animal	This is often overlooked
All	Manure	Feed intake per adult animal * dry matter digestibility of diet coefficient * population of adult animals	Alternatives include weighing the manure or if it is sold to ask how much was sold

Measurement method 1: Farmer reported production

Method of data collection and data needed

To calculate the production of meat and hides in a given time period a number of specific variables need to be measured, including the current number of animals, the number born, slaughtered or sold in that time period, and the total amount of products or services provided in that time period. If efficiency is being measured as well, the total amounts of all inputs provided should be collected.

In many countries, the Living Standard Measurement Studies (LSMS) carried out by the World Bank includes livestock production questions using recall from the past 12 months. Questions from the Tanzania

National Panel Survey (NPS) for 2010-2011 are illustrative of the wording and ordering of survey recall questions to estimate animal production.

The following questions were asked about each type of animal to obtain data on numbers owned, births, deaths, sales, and slaughtering (World Bank, 2017):

1. Did this household own any [ANIMAL] in the last 12 months? IF NOT INDICATED SEPARATELY, INCLUDE BABIES.
2. Number of [ANIMAL] owned 12 months ago.
3. How many [ANIMAL] does this household currently own? (Categorized as Indigenous, Improved Beef/Meat, Improved Dairy.)
4. How many [ANIMAL] were born in the past 12 months?
5. Have you bought any [ANIMAL] alive in the past 12 months?
6. How many [ANIMAL] have you bought alive in the last 12 months?
7. What was the total value of these purchases?
8. Did you receive any [ANIMAL] as a gift in the last 12 months?
9. How many [ANIMAL] did you receive as gifts in the last 12 months?
10. Have you lost any [ANIMAL] to DISEASE in the past 12 months?
11. How many [ANIMAL] have you lost to DISEASE in the past 12 months?
12. What was the value of these [ANIMAL]s lost to disease?
13. Have you lost any [ANIMAL] to THEFT in the past 12 months?
14. How many [ANIMAL] have you lost to THEFT in the past 12 months?
15. What was the value of these [ANIMAL]s lost to THEFT?
16. Have you sold any [ANIMAL] alive in the past 12 months?
17. How many [ANIMAL] have you sold alive in the past 12 months?
18. What was the total value of sales?
19. Did you slaughter any [ANIMAL] in the past 12 months?
20. How many [ANIMAL] did you slaughter in the past 12 months?
21. How many of the [ANIMAL] slaughtered did you sell?
22. What was the total value of the sold slaughtered [ANIMAL]?

These questions were asked for the following animal types: bulls, cows, steers, heifers, male calves, female calves, goats, sheep, pigs, chickens, turkeys, rabbits, donkeys, dogs, other___.

Another set of questions was asked about animal by-products:

1. Did your household produce any [PRODUCT] in the last 12 months?
2. During the last 12 months, for how many months did your household produce any [PRODUCT]?
3. During these months, what was the average quantity of [PRODUCT] produced per month? (unit in liters, kg or pieces)
4. Did you sell any of the [PRODUCT] that you produced in the last 12 months?
5. How much of the [PRODUCT] produced did you sell in the last 12 months?
6. What was the total value of sales of [PRODUCT] in the last 12 months?
7. Where did you sell most [PRODUCT] that you sold?
8. Who in your household decides what to do with these earnings? (up to 2 people)

These questions were asked for the following products: cow milk (traditional), cow milk (improved), chicken eggs (traditional), chicken eggs (improved), Ghee/Butter, Cheese/Yogurt, Honey, Skins and hides, Manure, Other____.

For the exact formatting of the questionnaire and the response categories see the NPS 2010-11 Agriculture Questionnaire (pp. 30-34).

Unit of analysis

The unit of analysis depends on the scale. Production can be measured per animal, per land area or per household. The unit of measure depends on the product (see Table 2).

Limitations regarding estimation and interpretation

Farmer recall may have large errors if the recall period is too long. The length of time that is reasonable depends on the livestock activity and its context. In many contexts, asking for recall over the past 12 months for births, sales, and slaughter of large animals (cattle, goats, pigs, sheep) may be reasonable. However, with poultry a 12-month period may have several cycles of birth and slaughter, so a monthly or quarterly recall may result in more precise information. Milk and egg production estimates may only be reliable for a week or a month. For many products, producers may not measure quantities precisely (e.g., kg of meat, liters of milk). In such cases, if precise production is needed, researchers may need to physically measure production.

Measurement method 2: Livestock models

Method of data collection and data needed

Process-based models of livestock production have been developed to enable estimating how changes in management practices will affect animal productivity. Jones et al. (2015) list some of the key dynamic livestock models as:

- Ruminant (Herrero et al., 2013, 1996)
- LiveSim (Rufino et al., 2009)
- CNCPS (Ruitz et al., 2002)
- Grazplan (Freer et al., 1997)
- GLEAM (Gerber et al., 2014)

Unit of analysis

The unit of analysis depends on the model and may be the individual animal, the herd or the farm.

Limitations regarding estimation and interpretation

The validity of these models depends on both the quality of the data used to set up the model and the appropriateness of the assumptions for the given model in the particular research context.

Metric 2: Rating of animal productivity

Description of the metric

Farmer ratings can be used to evaluate a new breed or how a management practice or technology has affected production, even without estimating the actual changes in production. These ratings can be carried out as a participatory exercise where farmers identify the important attributes and rate the alternatives by these attributes. Separating the analysis by type of farmer or by gender may provide useful information about how various groups perceive the alternatives and what attributes they prioritize.

Ratings (or scores) are generally easier to analyze statistically than rankings (see details in Measurement method 1: Farmer reported production).

Method of data collection and data needed

The following method to collect matrix rating data on animal productivity is an excerpt from “A guide to indicators and methods for assessing the contribution of livestock keeping to livelihoods of the poor” by Dorward et al. (2005).

Box 2. Species, function and alternatives matrix (Excerpted from Dorward et al. 2005, pp.20-22)

Purpose: to identify potential livestock species and livestock keeping activities of beneficiaries, the functions of such livestock keeping for each species, and alternative (non-livestock based) ways of achieving these functions

Contribution: Determination of beneficiaries' priorities & options in livestock keeping

Activity: Construct a matrix for ranking species according to the potential future importance of their contribution to different functions.

Species, future functions and non-livestock alternatives:

	Potential future importance in contribution to functional achievements			
Functions	Species 1	Species 2	Species 3	Alternatives (non-livestock)
F1	Rank or score	Rank or score	Rank or score	Specify & Rank or score
F2	Rank or score	Rank or score	Rank or score	Specify & Rank or score
F3	Rank or score	Rank or score	Rank or score	Specify & Rank or score

Method: This activity builds on activity 1 and takes it further by looking at the potential importance of livestock (and other activities) in achieving important functions in the future. The prioritisation of functions can be taken from matrix 2 (Functions, priorities & preliminary indicators) with appropriate discussion of the ways that functions' importance may change in the future. Then for each function the potential contribution of different species *and of alternative non-livestock keeping activities* should be ranked (or scored) with regard to their relative importance in fulfilling livelihood functions in the future.

Example: Species, future functions and non-livestock alternatives matrix for crop- livestock farmers, Yapacani, Bolivia.

	Potential future importance in contribution to functional achievements				
Functions	Chickens	Cows	Pigs	Ducks	Alternatives (non-livestock)
Consumption	1	0	3	2	Buying meat
Income	2	3	2	3	Selling labour
Buffering	1	3	2	1	Loans/ Credit/ Family
Accumulation	3	2	1	3	Investing in skills, Renting-in land
Insurance	0	1	2	0	Loans/ Credit/ Family
Social	0	1	2	3	Owning a car or motorbike

The process of discussion in order to complete this matrix should stimulate thought on the nature of change for both researchers and beneficiaries, which can be very useful. It is essential that reference is made to identified functions and the processes of change when discussing the changes in the relative importance of functions fulfilled by keeping various species. In our example chicken is becoming more important in consumption, probably due to its diminishing value due to the supply of cheap chickens nearby from a large chicken farm. Ducks' importance as providers of buffering is perceived as increasing, probably due to their higher relative value compared to chickens.

Unit of analysis

The unit of analysis is the individual respondent and his/her perception of the attributes of each alternative.

Limitations regarding estimation and interpretation

Farmer ratings provide only a coarse estimate of relative production of animals. Farmer ratings may not be easily compared across sites, especially if cultural factors vary, thus affecting the importance of various attributes.

Metric 3: Herd composition

Description of the metric

The composition of the herd can be indicative of its production objective and productivity. A study in Botswana (Behnke 1987) found that commercially-oriented herds had fewer oxen (which typically sold for meat) than herds used primarily for accumulation (whether as insurance or as a cultural priority). That study found herd composition to be a better indicator of commercialization than net sales because of the frequency of emergency sales when crops failed. In Western Kenya where the focus is on dairy production, cows form a large portion of the herd, whereas in the Ethiopian Rift Valley draft power is a key service so oxen are more numerous than cows (Baudron et al., 2014). A detailed study of Maasai herd composition (de Leeuw et al., 1991) showed that larger herds had higher heifer to cow ratios and calf to cow ratios, which indicates they are growing the fastest. Small herds either had lower reproductive success or owners were prioritizing selling young females over growing their herd.

Measurement method 1: Household animal inventory

Method of data collection and data needed

The following method to collect data on herd composition is an excerpt from "A guide to indicators and methods for assessing the contribution of livestock keeping to livelihoods of the poor" by Dorward et al. (2005).

Box 3. Household animal inventory (Excerpted from Dorward et al. 2005, p.19)

Purpose: to identify in more detail the herd or flock structure and composition for each livestock species and significant changes over the year, to gain greater understanding of livestock keeping activities. (Note: The collection of livestock numbers by species, sex and age provides data that can also be used to estimate a limited number of quantitative production parameters such as fertility rates and some information on age at first calving, lambing or sowing. There is also the possibility of making some very simple estimates on the value of investment in livestock and their generation of income. Simple production parameters can also be used in herd models to develop ideas on production and productivity (for further references see Dorward et al. 2005).

Contribution: Determination of beneficiaries' current assets, activities, priorities & options in livestock keeping

Activity: Construct a matrix showing the flock/ herd composition for each species kept. Ranges of numbers kept can be related to parts of the year to show seasonal variation (Matrix 3).

Matrix 3. Household inventory x season

	Species 1				Species 2				Species 3				Species …			
Season	a	b	c	D	a	b	c	d	a	b	c	d	a	b	c	d
Adult males																
Adult females,																
Young males																
Young females																
etc…																

Method: List the species kept by the household in columns and major classes in rows. In each cell note down the number of animals kept. To capture seasonal variation it may be necessary to draw up a separate table for different seasons, to note down in each cell particular seasonal events or changes, or to draw different sub-columns for each species to represent inventory changes between different times of year. The matrix layout above uses this approach, allowing for four different seasonal periods in the year.

Unit of analysis

The unit of analysis is the household/farm livestock.

Limitations regarding estimation and interpretation

This metric may be difficult to estimate at the community level when livestock ownership is concentrated among a few farmers. Interpreting the meaning of changes in herd composition requires in-depth knowledge of the context and driving forces behind farmers' behaviors. The data collected from the animal inventory can be complemented by qualitative information from interviews or focus group discussions inquiring about the uses of each type of animal and farmers' reasoning for changes in herd composition.

Metric 4: Net Commercial Offtake

Description of the metric

Net commercial offtake at the community (or wider) scale refers to net sales from that community or region, and focuses on the number of animals being bought and sold and not on animal by-products (like milk or eggs).

Measurement method 1: Household survey

Method of data collection and data needed

Data on sales and purchases from a household survey can be used to calculate net commercial offtake for a community or region, as described for Ethiopia in Negassa and Jabbar (2008). For details on survey questions, see the Animal Productivity indicator. The formula is:

$$\textit{Net commercial offtake} = (\textit{Total population} / \textit{sampled population}) * \Sigma\, (\textit{Sales} - \textit{Purchases}) / (0.5*(\textit{Opening stock} + \textit{Ending stock}))$$

where Σ represents the sum of all randomly sampled individuals surveyed for a given community or region. If a stratified sampling strategy is utilized, the sampling weights should be used instead of the first term (Total population / sampled population).

Unit of analysis

The unit of analysis is household.

Limitations regarding estimation and interpretation

Livestock ownership is not evenly distributed among households, so unless a multi-stage sampling design is used to target households that own livestock, standard errors may be high.

Measurement method 2: Regional sales figures

Method of data collection and data needed

It is often possible to obtain some estimate of the total number of animals imported or exported from an administrative region, such as a district or province. Government record keeping on the movement of livestock varies widely from one country to another. One promising example is the Tanzanian Ministry of Industry and Trade establishing market information systems in 53 markets across the country to collect price and quantities of animals sold (Mapunda et al., 2011).

Unit of analysis

The unit of analysis is region.

Limitations regarding estimation and interpretation

Regional figures may not capture all of the livestock trading being carried out.

Shortcomings regarding estimation and interpretation of the indicator

Smallholder farmers value livestock for uses beyond its productivity. In many contexts, farmers rely on livestock for savings, insurance, social prestige, and cultural functions (Dorward et al., 2005), which are beyond the production metrics listed. This needs to be considered when evaluating SI technologies. The savings and insurance uses relate to risk and reducing consumption variability, which in turn relates to the economic domain. The social prestige use of livestock could be detected through participatory wealth ranking (described in the economic domain). Cultural functions of livestock may be assessed through the social cohesion indicator in the social domain. Detailed methods for implementing a "Species by Function Matrix" and a "Function Priority and Preliminary Indicator Matrix" are provided in Dorward et al. (2005).

Variability of Productivity

Description of the indicator

The risk of low yields or low animal productivity may be even more important in some contexts than the average productivity. Quantifying the variability of productivity over time and space can be an important measure of this risk. Variability over space for a given time period and variability over time for a given environment are both important, but for different reasons. Variability over space has direct implications for recommendation domains. Yield stability across multiple sites is an important attribute for plant breeders seeking a variety that can perform well in many contexts. In contrast, variability over time for a given context is often driven by one or more variables, especially climate. Climate change adaptation has focused on identifying options less sensitive to poor climate years while still performing well on good and average years.

Method of data collection and data needed

There are many methods for assessing yield variability. The plant breeding literature has a long and rich literature for estimating yield stability as part of the analysis of researched varieties (e.g., Freeman 1973). Multiple regression makes it possible to analyze genotype by environment interactions to identify which factors contribute most to variability in yield. For example, Gupta and Ndoye (1990) analyzed pearl millet varieties in Senegal by comparing mean yield, the slope of the yield response to an environmental index, and the variability of that response.

The simplest method to assess changes in yield variability is to calculate the sample variance from a time series or cross-sectional data set. The square root of the variance is the standard deviation.

Unit of analysis

The bias adjusted sample variance may be calculated as:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

where n is the sample size, x_i is the *i*th observation, and $\bar{x}$ is the sample mean.

Limitations regarding estimation and interpretation

Simple comparison of the yield variance may be misleading if the driving factors are not controlled for. Analysis of variance and multiple linear regression are more robust methods that may be used.

Input Use Efficiency

Description of the indicator

The concept of efficiency focuses on avoiding wastage. Input efficiency is supposed to increase the performance of the system and minimize losses to the environment, for example if excess fertilizer like nitrogen ends up in a river or emitted to the environment. Finding avenues that ensure efficient use of applied nitrogen by the plants will be critical in avoiding wastage. Examining input use efficiency depends on the goal of the project and the biophysical, social, and economic context (Dobermann, 2007; Fixen et al., 2015). Input use efficiency is especially important to measure where a specific input is limiting (such as water in some contexts).

Method of data collection and data needed

For this metric, data on production (output), area of land used to produce this output, and inputs are required for this calculation. A main example of use of this metrics is in the use of chemical fertilizer, particularly nitrogen. Data on crop yield without use of input (nitrogen) and crop yield where input (Nitrogen) is applied are collected to compute this metric. Both partial factor productivity and agronomic efficiency metrics can be computed. The former examines 'productivity of the system relative to the input used' and the latter metric, 'productivity gained through the use of additional nutrients' (Fixen et al., 2015).

Unit of analysis

Partial Factor productivity is estimated as:

$$EPFP = Q/F$$

where Q is yield and F is the quantity of fertilizer applied to it during the season, and

$$EAgronomic = (Q - Q_0)/F$$

where Q_0 is yield obtained on the plot without fertilizer and Q is yield obtained from a plot where input F was used. The calculations on the metrics can be found in Fixen et al. (2015).

Limitations regarding estimation and interpretation

Interpretation of the input efficiency should be taken with caution because isolating a contribution to output involves controlling for other factors that might affect input use. For example, if the plot where fertilizer is applied receives optimal rainfall in the growing season, the complementarity with fertilizer use might increase the input use efficiency more than in an area where sub-optimal rainfall is received. Explaining the input use efficiency while also considering complementary conditions is important.

Yield Gap

Description of the indicator

The yield gap (Y_g) concept is based on definition and measurement of yield potential. Yield gap may be defined as the difference between yield potential or water-limited yield and actual yields (van Ittersum et al. 2013; Lobell, et al., 2009); however, Mueller et al. (2012) and Tittonell and Giller (2013) define yield gap as the difference between 'attainable yields' and landscape-level observed yields. Examining yield gaps is important as policy makers work toward ensuring food security at the micro and macro levels in a sustainable manner. Due to favorable climate, soil quality, and access to irrigation, some regions may have greater potential for sustainable agricultural intensification. To understand the concept of yield gap, we will define some key components in the calculation:

1) Yield potential (Y_p) is the yield of a crop cultivar grown with no limitations of water and nutrients, and biotic stress is effectively controlled (Evans, 1996; van Ittersum and Rabbinge, 1997). Therefore, for a given site when a crop is grown to achieve yield potential, growth rate is determined by solar radiation, temperature, water supply, and genetic traits that govern the length of the growing period. Yield potential is defined for irrigated systems where crops are given adequate water throughout the growth process.
2) Water-limited potential yield (Y_w) for rain-fed crops is similar to yield potential but crop growth is limited by water supply. It is therefore influenced by soil type (water holding capacity and rooting depth) and field topography (runoff) (van Ittersum et al., 2013).
3) Locally attainable yield (Y_l) is the maximum yield achievable by resources available to farmers in the most productive fields. These yields are more conservative than absolute biophysical potential yields because Y_l is achieved using current technology and management techniques (Tittonell and Giller, 2012; Mueller et al., 2012).
4) Average yield (Y_a) is the actual yield achieved in a farmer's field.

Yield gap can be estimated as the difference between and Y_a and Y_p, Y_w or Y_l. Lobell et al. (2009) identify four methods of estimating yield gap at local levels that can be used to obtain the yield potential (maximum attainable yield). These include: 1) crop model simulations, 2) maximum farmers' yield based on surveys, 3) yield contest, and 4) field experiments. In this case we assume yield ceiling is obtained from crop modelling or field experiments.

Yield potential is estimated mainly using crop models that assume perfect management using lack of all yield-reducing factors. A shortcoming of crop models is they lack the sensitivity to short-term abiotic stress which leads to higher estimates of potential yields than would occur in the field. As for field experiments, the difficulty of achieving perfect yield conditions increases with plot size and year-to-year climate variation (which may be considerable at any given location), therefore requiring a time series of experimental data to ensure a mean estimate that reflects a range of climates. A combination of crop model simulation and field experiments is recommended to provide more robust estimates. The use of maximum farmer yields may be used to estimate yield potential in locations where farmers intensively manage a crop with the possibility of achieving the yield potential. Actual yield estimates in the field (i.e., crop cuts) are recommended to complement farmer reported values (Lobell, et al., 2009). Crop production capacity under rain-fed and irrigated conditions can be analyzed as a benchmark by estimating yield potential and water-limited yields. This is essential for sustainable agricultural intensification.

Method of data collection and data needed

To estimate yield potential, data on the actual yield of the farmer or scientist should be collected. This can be done utilizing the methods used to estimate "Yield" in the Crop Productivity Indicator. Yield estimates can be obtained either through farmer recall or through actual measurements (e.g., crop cuts) of experimental or farmer-managed plots. In this example, we assume that to obtain the yield potential, the scientists used crop-modelled yield potential that is calibrated for the given cultivar and biophysical conditions of the area. For the data needed to calibrate a model, such as Decision Support System Agrotechnology Transfer (DSSAT) consult Jones et al. (2003) and for Agricultural Production System sIMulator (APSIM) consult Keating et al. (2003). One may use the help of a crop modeler to obtain these estimates or obtain these data through available secondary sources. See also the crop modeling method for yield estimates previously detailed.

Unit of analysis

Once these data have been obtained, we can calculate the yield gap in kilograms per hectare. Where crop modelling is used to determine the potential yield, there is need to adjust this yield to reflect the 'actual' farm conditions. For example, at the field experiment or in the crop model, there is timely application and management of inputs which may not be the case in most farmers' fields. Lobell et al. (2009), have indicated in their study that there is no area that has achieved a yield above 80% of the modelled yield potential, and we suggest that in most studies an adjustment factor should be used. The yield gap (kg/ha) is calculated as:

$$y_g = (y_p * k) - y_a$$

where y_p is the yield potential, y_a is actual yield, and k is an adjustment factor.

Limitations regarding estimation and interpretation

Yield gaps tend to be difficult to compare across location and studies because of inconsistent terminology, concepts, and methods. Measured yields at experimental fields may be biased because stations are often situated on soils with suitable topography, making it poorly representative of surrounding topography. Yield potential requires perfection in the management of all other determining factors, from sowing (plant population, supply and balance of 17 essential nutrients, and protection against loses from insects, weeds, and disease) to maturity.

Cropping Intensity

Description of the indicator

Cropping intensity is defined as the number of crops a farmer grows in a given agricultural year on the same field (Raut et al., 2011), and is another means for intensification of production from the same plot of land. The indicator focuses on the agricultural year and assesses the number of crops grown. In regions such as Sub-Saharan Africa where there is minimal irrigation, improvement in the number of crops grown may be an intensification strategy for households.

Method of data collection and data needed

Data for this metric are collected via survey. Agricultural surveys, such as those used by the World Bank (NBS, 2014), divide the agricultural year into two seasons. Crop production is enumerated for both seasons. Therefore, data for this metric should be collected for both seasons. Data should be collected on whether the crop was grown in the season. In the LSMS-ISA Surveys, the household is asked which crops they grew in each season, area planted, and the amount that was harvested.

Unit of analysis

For survey data, the cropping intensity may range from 0 to 2. If the household grows a given crop in both seasons, then the count is two crops; and if the household does not grow a crop during the agricultural year, then the value is 0.

Limitations regarding estimation and interpretation

The cropping intensity can have a higher maximum value than a score of 2 where early maturing crop varieties are used and more than two crops can be grown or where technology like irrigation is used. At the field level, these surveys should be modified to ensure that changes in technology that increase cropping intensity are captured.

References

Anderson, J. and Ingram, J. 1993.Tropical Soil Biology and Fertility – A Methods Manual 2nd edition, CABI.

Baker, D. and Pica-Ciamarra, U. 2010. Routine livestock data in sub-Saharan Africa: collection and use. FAO Publication: Rome.

Baudron, F., Jaleta, M., Okitoi, O. and Tegegn, A., 2014. Conservation agriculture in African mixed crop-livestock systems: Expanding the niche. Agriculture, Ecosystems & Environment, 187, pp.171-182.

Behnke, R.H. 1987. Cattle accumulation and the commercialization of the traditional livestock industry in Botswana. *Agricultural Systems*, *24*(1), pp.1-29.

Coe, R. 2002. Analyzing Ranking and Rating Data from Participatory On-Farm Trials. p.44 in Bellon, M.R. and Reeves, J., *Quantitative analysis of data from participatory methods in plant breeding*. CIMMYT. http://libcatalog.cimmyt.org/download/cim/448231.pdf

De Groote, H., Rutto, E., Odhiambo, G., Kanampiu, F., Khan, Z., Coe, R. and Vanlauwe, B. 2010. Participatory evaluation of integrated pest and soil fertility management options using ordered categorical data analysis. *Agricultural Systems*, 103(5), pp.233-244.

Dobermann, A. 2007. Nutrient use efficiency–measurement and management. In: *Fertilizer best management practices*, Papers presented at the IFA International Workshop on Fertilizer Best Management Practices 7-9 March 2007, Brussels, Belgium, International Fertilizer Industry Association, Paris, France.

Dorward, A., Anderson, S., Paz, R., Pattison, J., Vera, E. S., Nava, Y. & Rushton, J. 2005. A guide to indicators & methods for assessing the contribution of livestock keeping to livelihoods of the poor. Department of Agricultural Sciences, Imperial College London Publication: Kent, United Kingdom.

Evans, L. T. 1996. Crop evolution, adaptation and yield. Cambridge University Press

Fixen, P., Brentrup, F., Bruulsema, T.W., Garcia, F., Norton, R. and Zingore, S. 2015. Nutrient/fertilizer use efficiency: measurement, current situation and trends. Managing Water and Fertilizer for Sustainable Agricultural Intensification, p.8.

Fermont, A. and Benson, T. 2011. Estimating yield of food crops grown by smallholder farmers. *International Food Policy Research Institute, Washington D.C.*, pp.1-68.

Freeman, Gr H. "Statistical methods for the analysis of genotype-environment interactions." *Heredity* 31.3 (1973): 339-354.

Holzworth, D.P., Huth, N.I., Zurcher, E.J., Herrmann, N.I., McLean, G., Chenu, K., van Oosterom, E.J., Snow, V., Murphy, C., Moore, A.D. and Brown, H., 2014. APSIM–evolution towards a new generation of

Jones, J. W., J. M. Antle, B. O. Basso, K. J. Boote, R. T. Conant, I. Foster, H. C. J. Godfray, M. Herrero, R. E. Howitt, S. Janssen, B. A. Keating, R. Munoz-Carpena, C. H. Porter, C. Rosenzweig, and T. R. Wheeler. 2015. Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: State of Agricultural Systems Science. AgMIP, <http://goo.gl/f4eVl4>.

ISPC. 2014. Data, metrics and monitoring in CGIAR – a strategic study. Rome, Italy, CGIAR Independent Science and Partnership Council (ISPC). 88 pp.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W. 2003. The DSSAT cropping system model. *European Journal of Agronomy* 18, 235–265.

Kawano, K., 1990. Harvest index and evaluation of major food crop cultivars in the tropics. *Euphytica*, 46(3), pp.195-202.

Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D., Huth, N.I., Hargreaves, J.N., Meinke, H., Hochman, Z. and McLean, G. 2003. An overview of APSIM, a model designed for farming systems simulation. *European journal of agronomy*, 18(3), pp.267-288.

de Leeuw, P N; P P Semenye; C P Peacock and B E Grandin. 1991. Productivity of Cattle and Smallstock. Ch. 7 in: Solomon Bekure, de Leeuw P N. Grandin B E and Neate P J H (eds). *Maasai herding: An analysis of the livestock production system of Maasai pastoralists in eastern Kajiado District, Kenya.* ILCA Systems Study 4. ILCA (International Livestock Centre for Africa), Addis Ababa, Ethiopia. 172 pp. http://www.fao.org/wairdocs/ILRI/x5552E/x5552e00.htm

Lesnoff, M., Lancelot, R., Moulin, C.-H., Messad, S., Juanès, X. & Sahut, C. 2011. Calculation of demographic parameters in tropical livestock herds – a discrete time approach with LASER animal-based monitoring data. Editions Quae, CIRAD: Montpellier, France.

Lobell, D. B. The use of satellite data for crop yield gap analysis. *F. Crop. Res.* 143, 56–64 (2013).

Lobell, D. B., Cassman, K. G., and Field, C. B. 2009. Crop yield gaps: their importance, magnitudes, and causes. *Annual Review of Environment and Resources*, 34 (1), 179.

Mapunda, A., Chassama, J. and Pica-Ciamarra, U. 2011. Livestock market data: Collection, dissemination and use in Tanzania. Livestock Data Innovation in Africa Brief 7. Rome, Italy: FAO.

Marten, G. C, J. S. Shenk, and F. E. Barton II, editors. 1989. *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality.* U.S. Department of Agriculture, Agriculture Handbook No. 643 (revised with supplements), 110 p.

Mead, R. and Willey, R.W., 1980. The concept of a 'land equivalent ratio' and advantages in yields from intercropping. *Experimental Agriculture*, 16(03), pp.217-228.

Messina, J., Peter, B., and Snapp, S. Re-evaluating the Malawian Farm Input Subsidy Programme. *N Plants* Accepted (2017).

Moss, J., Morley, P., Baker, D., Al-Moadhen, H., Downie, R. 2016 Improving methods for estimating livestock production and productivity – literature review. Improving agricultural and rural statistics Technical Report Series GO-11-2016.

Mourice, S. K. et al. 2014. Maize cultivar specific parameters for decision support system for agrotechnology transfer (DSSAT) application in Tanzania. *American Journal of Plant Sciences* 5.6: 821.

Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N., & Foley, J. A. (2012). Closing yield gaps through nutrient and water management. *Nature*, 490 (7419), 254-257.

National Bureau of Statistics, United Republic of Tanzania, National Panel Survey 2010-2011, Agriculture questionnaire. Available from: http://go.worldbank.org/EJMAC1YDY0

Negassa, A. and Jabbar, M., 2008. Livestock ownership, commercial off-take rates and their determinants in Ethiopia. ILRI (aka ILCA and ILRAD).

Newbould, P. J. (1967) Methods of Estimating the Primary Production of Forests. Blackwell Scientific Publications, Oxford.

Palm, C.A., Gachengo, C.N., Delve, R.J., Cadisch, G. and Giller, K.E., 2001. Organic inputs for soil fertility management in tropical agroecosystems: application of an organic resource database. *Agriculture, Ecosystems & Environment*, 83(1), pp.27-42.

Peng, D., Huang, J., Li., C., Liu, L., Huang., W., Wang, F. and Yang., X. Modelling paddy rice yield using MODIS data. *Agric. For. Meteorol.* 184, 107–116 (2014).

Pica-Ciamarra, U., Baker, D., Morgan, N., Zezza, A., Azzarri, C., Ly, C., Nsiima, L., Nouala, S., Okello, P. and Sserugga, J. 2014. *Investing in the livestock sector - why good numbers matter: a sourcebook for decision makers on how to improve livestock data.* International Bank for Reconstruction and Development/The World Bank Publication: Washington D.C.

Proctor, J. (1983) Tropical Forest Litterfall 1. Problems of data comparison. In: Sutton, S.L., Whitmore, T.C. and Chardwick, A.C. (eds.), Tropical Rain Forest: Ecology and Management. Blackwell Scientific Publications, Oxford, pp.267-273

Reeves, M. C., Zhao, M. and Running, S. W. Usefulness and limits of MODIS GPP for estimating wheat yield. *Int. J. Remote Sens.* 26**,** 1403–1421 (2005).

Sapkota, T.B., Jat, M.L., Jat, R.K., Kapoor, P. and Stirling, C., 2016. Yield Estimation of Food and Non-Food Crops in Smallholder Production Systems. In: Rosenstock, T.S., Rufino, M.C., Butterbach-Bahl, K., Wollenberg, E. and Richards, M., (Eds.) *Methods for Measuring Greenhouse Gas Balances and Evaluating Mitigation Options in Smallholder Agriculture.* Springer

Raut, N., Sitaula, B.K., Vatn, A. and Paudel, G.S., 2011. Determinants of adoption and extent of agricultural intensification in the central mid-hills of Nepal. Journal of Sustainable Development, 4(4), p.47.

Stuth, J., Jama, A., Tolleson, D. 2003. Direct and indirect means of predicting forage quality through near infrared reflectance spectroscopy. *Field Crops Research* 84, 45–56

Tittonell, P., and Giller, K. E. (2013). When yield gaps are poverty traps: The paradigm of ecological intensification in African smallholder agriculture. *Field Crops Research*, 143, 76-90.

Van Ittersum, M. K., and Rabbinge, R. (1997). Concepts in production ecology for analysis and quantification of agricultural input-output combinations. *Field Crops Research*, 52 (3), 197-208.

van Ittersum, M. K., Cassman, K. G., Grassini, P., Wolf, J., Tittonell, P., & Hochman, Z. (2013). Yield gap analysis with local to global relevance—a review. *Field Crops Research*, 143, 4-17.

World Bank, FAO and UN. 2010. Global strategy to improve agricultural and rural statistics. Report number 56719-GLB. IBRD/World Bank Publication: Washington, D.C.

World Bank 2017. Living Standard Measurement Study Survey Integrated Surveys on Agriculture (LSMS-ISA).

You, L., Ringler, C., Wood-Sichra, U., Robertson, R., Wood, S., Zhu, T., Nelson, G., Guo, Z. and Sun, Y. 2011. What is the irrigation potential for Africa? A combined biophysical and socioeconomic approach. *Food Policy*, 36 (6), 770-782.

Economic Domain (Note: The superscript letters (a,b,c) after each metric refer to the measurement methods in the right-hand column)

ECONOMIC DOMAIN					
Indicator	**Field/plot level metrics**	**Farm level metrics**	**Household level metrics**	**Community/ Landscape + metrics**	**Measurement method**
Profitability	Net income [a] ($/crop/ha/season) Gross margin [a]	Net income [a,c] (total net income for all farm activities) Gross margin [a]	Net income [a,c] (total net income for all farm activities)	Contribution to regional or national GDP [b]	[a] Survey [b] Regional and national statistics [c] Participatory evaluation
Variability of profitability	Coefficient of variability of net income [a] Probability of low profitability [a,b]	Coefficient of variability of net income [a] Probability of low profitability [a,b]	Coefficient of variability of net income [a] Probability of low profitability [a,b]		[a] Survey [b] Farmer evaluation
Income diversification	N/A	Diversification index [a]	Diversification index [a] Number of income sources[a]		[a] Survey
Returns to land, labor, and inputs	Returns [a] (monetary value of output/input used)	Returns [a] (monetary value of output/input used)	Returns [a] (monetary value of output/input)		[a] Survey and productivity measurements
Input use intensity	Input per ha [a]	Input per ha [a]	Input per ha [a]		[a] Survey
Labor requirement	Labor requirement (hours/ha) [a,b] Farmer rating of labor [c]	Labor requirement (hours/ha) [a,b] Farmer rating of labor [c]	Labor requirement (hours) [a,b] Farmer rating of labor [c]		[a] Recall survey [b] Direct observation [c] Farmer evaluation
Poverty	N/A	N/A	Asset index [a] Per capita hh consumption expenditure [a] Wealth categorization [b]	Poverty headcount ratio [a] Asset wealth categorization [b]	[a] Survey [b] Participatory exercise
Market participation	N/A	N/A	% production sold [a]	Total sales [a]	[a] Survey
Market orientation	N/A	N/A	% land in cash crops [a] Market orientation index [a]		[a] Survey

Economic Domain

Profitability

Description of the indicator

Profitability measures the viability of the technology using its revenues and expenses (Hofstrand, 2009). This analysis may be done over a season, year or multiple years depending on the objective of the research and the technology being assessed.

In instances where new technology is provided to a farmer, the farmer will have to consider the costs of changing from one practice to another and the economic returns resulting from that change. The farmer may assess this in terms of grain produced, labor allocated or time spent producing output using the new technology compared to the conventional technology (CIMMYT, 1988). Therefore, in assessing the profitability of a new technology, it is recommended that the scientist performs a budgeting analysis that lists the quantities of inputs used in the production process and the output generated (Amir and Knipscheer, 1989). This budgeting exercise will be used to assess the monetary value of inputs and outputs. It is important to note that in these economic analyses, farmers have limited or scarce resources and there is an opportunity cost for the choice they make. For example, the opportunity cost of allocating more labor to the new technology may be the foregone income from working off-farm or leisure time. There are a number of profitability metrics that have been used in on-farm agricultural research to evaluate technologies, including partial budget analysis, gross-margin, net income, break-even analysis, and benefit cost ratio. Partial budget is used in minor changes in farming practice and may be used for a single intervention. Gross margin is used to calculate profitability of alternative plans and estimates the returns to a technology above the variable costs (Rural Solutions, 2012; Amir and Knipscheer, 1989). Net income examines returns to a technology but includes the fixed or overhead costs of production -- unlike gross margins -- but may also be cumbersome to calculate. Break even analysis is important in examining the point at which returns equal the costs, and a breakeven price can be calculated for a technology that provides an indication of price at which the output can be sold to recover the costs (Dillon and Hardaker, 1980 Amir and Knipscheer, 1989). Benefit cost analysis (BCA) is similar to gross margin except that it accounts for the non-cash costs and benefits (Dillon and Hardaker, 1980. The following sections with focus on discussing the estimation of the net income and gross margin indicator. For more information on other metrics on profitability, please refer to Dillon and Hardaker (1980), Amir and Knipscheer (1989), and CIMMYT (1988).

Metric 1: Net income

Description of the metric

Net income metric is derived from incomes and costs of production (variable and fixed or overhead costs). Costs of production includes variable and fixed costs, such as labor, fertilizers, seed, and feed, plus non-cash expenses that might include annual depreciation of equipment, such as a tractor. Depreciation is added to the net income calculation because for long-term profitability, the farmer must be able to replace the equipment once it wears out. In addition, interest paid on loans for the business should be added to the expenses (costs). When considering the decision to adopt a technology, a farmer examines whether there are additional net gains from the technology or innovation (for brevity innovation and technology will be used interchangeably).

Method of data collection and data needed

Data used to calculate net income of a technology are usually collected by survey. Recall surveys or diaries are used to enter data on the input costs, management, and income from output. Cost (price per unit of input) and quantity of inputs used in the production process are critical data. Standardized units such as kilograms or liters should be used. The approach to calculating net income at the field level is described below with the unit of analysis.

From the survey, it is recommended (but not limited) to collect the following types of data:

1) Quantity of inputs used in production (inorganic fertilizer, pesticides, labor, feed);
2) Price of each input above (used to calculate the cost of each input);
3) Other costs (cost of veterinary services, transportation costs, cost of vaccines, breeding fees);
4) Quantity and price for output (for example; price of milk and amount produced); and
5) Fixed costs that may include
    a. Housing (farmhouse for animal or crop production)
    b. Land rent
    c. Beginning stock (for animal production)
    d. Depreciation
    e. Taxes or interest

Standard surveys may not collect all this information but it is advisable to adjust the survey tool to meet the needs for calculating the metric. In Table 3 we provide an example in which the net income metric is calculated for the new maize technology.

Table 3. Example of single season activities for two technologies at the plot level

	Traditional Maize				Improved Maize Seeds			
Activity	*Labor Days	Type	Unit	Cost ($)	*Labor Days	Type	Unit	Cost ($)
Land Preparation	12				8			
Planting	7	seed	1kg	2	7	Seed	1kg	2
Fertilization (organic)	3	various	6kg	12	3	Various	6kg	12
Weeding	18				18			
Harvesting	8				8			
Land (rental)	0	land	hectare	0	0	Land	hectare	0
Chemical fertilizer	0		kg	0	3	DAP	50Kg	20
Chemical fertilizer					3	Urea	50kg	18
Pesticide application					4	TKC	3Lt	8
Farm machinery (tractor)	0				3			
Total	48			14	57			60
*Labor day is 6 hours per day at a rate of $4. Wage rate is obtained from hired wage rate in surveys (Table 2)								

(Adapted from Avila, 2016: http://fisheries.tamu.edu/files/2013/09/SRAC-Publication-No.-4402-Determining-the-Profitability-of-an-Aquaculture-Business-Using-Income-Statements-and-Enterprise-Budgets.pdf.).

In Table 3, the costs of production for two technologies, traditional maize and improved maize, are listed. These can be collected from an Agricultural Survey. It is important to note the cost of labor is an important input for this calculation and enumerating the wage rate of labor is important. For small plot experiments where the farmer works on his own plot, the wage rate may be difficult to obtain but these data can be inferred from the going daily wage for hired labor in that area. For example, if the household hires an extra laborer at $4 per day, this may be used as the going wage rate. In comprehensive surveys, such as the LSMS-ISA surveys, the labor section asks about wage rate for hired labor. Where this wage rate is noted, this amount can be used. In our example, a wage rate of $4 is the going rate[1]. The total cost of inputs other than labor are calculated (Table 1) from the data collected in the survey. Please note that total area (in acres or hectares) to which this technology is applied is important, as well as the total cost of each input used.

[1] The labor wage rate in most developing countries may be difficult to assess due to lack of quality data or missing labor markets. Caution should be taken to ensure that enumeration of the wage rate does not under- or over-estimate the returns to labor.

Table 4. Comparative analysis on profitability of two technologies (plot level and single season)

Criteria	Traditional Maize	Improved maize
Production (kg)		
Maize	1300	2800
Yield (kg/ha)		
Maize	1300	2800
Gross Income ($)		
Maize	260	560
Total Income	260	560
Variable Costs ($)		
Labor ($)	192	228
Other inputs (Cost $)	12	60
Fixed costs: ($)		
Land	20	20
Depreciation of tractor	0	12
Total Costs	224	320
Profitability Indicators ($)		
Net Income/ha	12	240
Net returns/labor day	4	8

In Table 4 the total production and value of production (in $ given the price of $0.2 per kg) are calculated. For simplicity, the area planted was one hectare for each technology. The variable costs are costs of labor and other commercial inputs, such as chemical fertilizers or pesticides (see Table 2 in Appendix). The fixed costs are the costs of land and depreciation of the tractor (e.g., $12for a tractor costing $120 to be used for 10 years with no ($0) salvage value). Net income is calculated as the difference between the total income and total costs (see equation 1) and the returns to labor are calculated using equation (2) (Avila, 2016; Engle, 2012).

Standard survey questions used to collect such data can be found in the Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) surveys sections 3,4,5, 10 and 11 (http://siteresources.worldbank.org/INTLSMS/Resources/3358986-1233781970982/5800988-1286190918867/NPS_Agriculture_Qx_English_(Year2)_v2.pdf).

Unit of analysis

The unit of measurement is the net income per hectare. For data collected during a season or agricultural year (not multiple years), we assume the net income should be examined as

$$Net\ income\ per\ hectare = Gross\ income - (Total\ costs)/total\ \ area\ (Ha).\ (1)$$

Gross income is the value of the produce sold (quantity sold *price) and total cost is the sum of the variable costs and fixed or overhead costs (cost of labor + seeds + chemical fertilizer + Land + depreciation of tractor…).

This may be used to compare between two technologies to assess which one performs better than the other.

Limitations regarding estimation and interpretation

This indicator has a number of limitations. Estimating net income requires computation of overhead or fixed costs. In cases of technology investments that cover more than one year, estimating these calculations may require some economic or accounting knowledge. This measure also requires more data than the gross margin metric, which may limit its usage where data are limited. Also, unlike benefit cost analysis, some of the non-cash benefits might not be considered in this analysis (the benefit cost analysis is recommended in cases where these are needed). Measuring profitability using the net income metric may require estimation of the opportunity cost of labor or the wage rates of labor which might be difficult to obtain at field level or in areas where labor markets are missing or incomplete Additionally, depending on the time horizon of the project, there is need to incorporate sensitivity analysis of the project given various price scenarios. Consideration of the internal rate of return for the project may be useful. These measures might require more expertise but are essential in explaining potential long-term adoption of the technology. Because of the limitations in assessing profitability of a given technology across scale, alternative measurement methods, such as gross margin or participatory evaluation, may be considered.

Metric 2: Gross margin

Description of the metric

Gross margin analysis is an alternative technique that can be used to calculate the profitability of technology. The measure focuses on variable costs and ignores the fixed or overhead costs since fixed costs will be incurred irrespective of the technology or level undertaken (Rural Solutions, 2012). Farmers do make decisions on what kind of technology may be economically efficient and technically feasible. Gross margin indicator may help in assessing technologies being developed for potential dissemination by comparing their benefits with conventional technologies . By focusing on the variable costs , the indicator is not data-intensive and may be simpler to compute than net income. Gross margin of the farm can also be calculated by summing up the gross margins on field activities.

Method of data collection and data needed

To calculate gross margin, the researcher must be aware of the farmer's objective for the project and account for the factors that will change or remain the same with the new technology. In this way, data on the costs that may vary in the production process are captured for analysis. Once these two aspects are considered, then data should be collected. There is no generic list on the variable costs but for the gross incomes, data on the price of output (crops or livestock products and by-products) plus the total production on the farm, should be collected. These data are used to calculate the value of the field or farm products (Gross Income). Data on variable costs should be collected and may include labor , feed, pesticides or herbicides, animal medicine, transportation, seeds (germplasm). These costs should be considered in terms of the new technology. These data may be collected using a survey that has been adapted for the technology. In cases of field experiments, journal entries should be made for the input and output costs . Price of the output can be obtained by performing a market survey on the price of the commodity. It is important to note that for field experiments, there is a difference between market cost and the "field price" (CIMMYT, 1988). In some cases, the market price may differ from the price that farmers receive because the product may be bought at the farm gate. In addition, there is seasonality of price, which may differ at the time of harvest and before harvest. These factors should be taken into consideration when assessing the price of the output.

Unit of analysis

Gross margin is calculated as:

$$GROSS\ INCOME - VARIABLE\ COSTS = GROSS\ MARGIN^{2}$$

At the farm or household level where there are multiple activities then total gross margin is calculated as

$$TOTAL\ GROSS\ MARGIN = \sum_{i=1}^{n} GROSS\ MARGIN_i$$

where i is the activity on the farm (e.g., crop, poultry, or ruminant production;(Amir and Knipscheer, 1989). Gross margin has an advantage in that it can be used to simply rank technologies while focusing on their profitability.

For illustration, suppose there is a new type of herbicide that is going to be tested on fields that use two different seeding rates to examine which one may provide higher benefits. Table 5 provides that illustration.

[2] Net Income is related to gross margin as follows: $NET\ \ INCOME = GROSS\ MARGIN - OVERHEAD\ or\ FIXED\ COSTS$

Table 5. Gross Margin for weed control technology

	Treatments*	
	1	2
Yield (kg/Ha)	1600	2000
Price ($/kg)	9	9
Gross income ($)	14400	18000
Cost of herbicide ($/Ha)	0	800
Cost labor to apply herbicides	0	600
Cost of labor for water haul ($/ha)	0	150
Cost of seeds	2500	2500
Cost of rental sprayer	0	150
Cost of labor for weeding	100	0
Variable costs	2600	4200
Gross Margin	11800	13800
* Note: 1. No weed control; 2. Herbicide use.		
Adapted from CYMMYT, 1988		

From the example above, treatment 2 may be chosen because it has a higher gross margin. But to complement this analysis, the researcher may also consider examining the breakeven price and breakeven output of the two technologies (Amir and Knipscheer, 1989). ***Gross margin of technologies that have different fixed costs should not be compared.*** It is important to also note that an intervention that has the highest gross margin may be more sensitive to price (input or output) variation. This risk may be calculated by comparing gross margin with different prices (Rural Solutions, 2012).

Limitations regarding estimation and interpretation

Gross margin should be interpreted with caution since it may not consider the fixed or overhead cost of production. Therefore, inferring that farm profits may be increased by scaling up production may not be realistic (Amir and Knipscheer, 1989; Dillon and Hardaker, 1980. In cases where data are collected at experimental fields at the station, production values may overestimate what farmers may produce on their actual fields. Gross margin should be complemented with additional analyses (e.g., break-even). This may provide information on break-even price--the level at which gains equal costs.

Metric 3: Participatory evaluation

Description of the metric

In evaluating technologies for agricultural research, administering a complete household and agricultural survey may not be an effective approach because some of the interventions are done on small plots with few participants. Even where survey data on profitability exist, participatory evaluation may provide additional information on farmers' perceptions, which then may be used to validate the survey data output or make comparisons.

Method of data collection and data needed

Soliciting information from farmers on the potential of the technology is required and this may be accomplished through a focus group or at an individual level. Data from this assessment will include examining or eliciting the following information:

1) Farmer's preference between the local (current) and new technology. This may be done by ranking or rating the technologies.
2) Whether farmers think the technology might be profitable if adopted. This must be done while considering the costs of inputs plus ability to market output from the new technology. The farmer indicates whether the new technology will be more profitable compared to the local technology, considering additional costs and benefits.
3) The potential for the farmer to allocate land on his/her farm to this technology and, if so, what percentage may be allocated to the new technology.
4) Whether the farmer intends to sell the output from the new technology to the market or use if for home consumption and, if so, what proportion?

From these questions, the farmers may be asked to rate the technology in terms of profitability. These ratings for example may be 1=low profitability, 2= intermediate, and 3=high profitability. In some cases, where the criteria for a technology are being assessed after trial, some of the reported criteria can be mapped to the profitability indicator. For example, farmers may indicate that the technology is "produced with little money compared to local technology" (Bellon, 2001).

Unit of analysis

Unit of analysis will be the average rating of the profitability of technology from farmers participating in the evaluation. If there is a high value, then the technology is profitable. Where a large number of households are randomly selected to participate in the evaluation, the distribution of the measure of profitability may be estimated for inference. Tests such as the Kruskal Wallis test may be used to determine whether the distribution functions across groups (wealth groups) are identical.

Limitations regarding estimation and interpretation

Using participatory evaluations may require training of scientists on how to select the population of study and collect data. This may not be a limitation but it may help the researcher to establish any additional value using this approach may provide, especially as it relates to the input cost and value of output to compute gross margin or net income.

Shortcomings regarding estimation and interpretation of the indicator

Estimation of profit and gross margin requires proper bookkeeping to ensure that all costs or required variable costs are captured. We recommend that required expertise from an economist should be sought when choosing between the two measures of profitability.

Variability of profitability

Description of the indicator

Variability of profit is an important metric because it provides a measure of variation from the mean that can be attributed from either the production or consumption side. This assessment may provide information beyond the average in assessing the profitability of the technology.

Metric 1: Coefficient of variation of net income

Description of the metric

The coefficient of variation (CV) is used to measure variability of a metric. CV is popular because it can be used to compare distributions with different units of measurement.

Method of data collection and data needed

In this case we will use CV to assess the variability of net income or gross margin. If a number of technologies are tested for profitability in different locations or multiple interventions on plots for a single technology, then net incomes or gross margins need to be calculated using the approach discussed earlier. Once these data on the net income or gross margin are obtained, the coefficient of variation can be calculated.

Unit of analysis

The CV is unitless but takes on values between zero and $\sqrt{N-1}$, where N is the number of non-negative values in the sample. The CV can then be computed as; $CV_P = \frac{\sigma_P}{\mu_p}$ where σ is the standard deviation of profitability measure and μ is the mean profitability of technologies tested. For example, if the CV of treatment 1 is higher than that of treatment 2, then the profit metric for treatment 1 is more variable than that of treatment 2.

Limitations regarding estimation and interpretation

Coefficient of variation is only meaningful when there are non-negative profits in the sample. Negative values might be possible due to crop failure or high costs of inputs. Caution must be taken in calculating a CV where negative values are observed because the measure is not valid for cross-item measurement unless other statistical measures are taken.

Variability of profitability is best measured directly through farmers' actual production, input costs, and output prices. However, it is also feasible to estimate variability in profits using production variability and price volatility values. Assumptions would need to be made about when farmers sell if their output prices vary seasonally (such as for maize in southern Africa). Thresholds and critical values for variability may relate to food security, poverty lines or similar objectives. Resilience of profitability to shocks (e.g., production, price, political) should be considered as a separate indicator.

Income Diversification

Description of the indicator

Income diversification [3] examines the number of income sources and the share that each income source contributes to total income. Income diversification has been observed as a livelihood strategy that may be due to a number of reasons:

1) household may diversify as a risk-coping strategy when income from one source is highly variable,
2) households may have high labor endowment compared to farmable land and the marginal productivity labor may be close to zero, hence the need to allocate family labor to other activities,
3) movement to higher-value crops due to market access and sufficient food for household consumption, and

[3] Income diversification is not synonymous with livelihood diversification.

4) economies of scope (Minot et al., 2006; Ersado, 2003).

Metric 1: Income diversification index

Description of the metric

The income diversification index literature contains a number of measures and indices, including the inverse of the Herfindhal Index (Ersado, 2006), the Herfindhal diversity Index (Block and Webb, 2001), the Simpson Index of diversity (Minot et al., 2006) and the number of income sources (NYS) (Ersado, 2006). Calculation of these indices requires data on agricultural activities (e.g., livestock production, cash crop production, food crop production), as well as off-farm work (e.g., wage employment, non-labor income such as remittances, self-employment in business or enterprise, and self-employment in agriculture). The calculated income diversification index may be used further to examine the factors affecting income diversification, which may include climate, asset wealth, or household and location characteristics.

Method of data collection and data needed

Estimating the income diversification index requires collecting data on the different income sources of the household. These data are usually collected via survey. In the household section, questions are asked for each member of the household to determine:

1) If they worked as a non-agriculture employee
2) If they worked in a self-owned business enterprise
3) If they worked in a salaried job
    a. Furthermore, a question on how much was earned from the job for that duration is asked to determine the total wage earned.

In addition, for the agricultural survey, questions on the agricultural activities in which a household engaged are enumerated. These include (World Bank, 2017):

1) Livestock production and the value of output,
2) Food crop production and its value, and
3) Cash crop production and value (a cash crop may be identified from the list of "traditional cash crops" due to the sales volume).

Examples of surveys where these data are enumerated are the LSMS surveys (NBS, 2014). Ersado (2006 examines income diversification and provides a categorization of off-farm income sources. This example is presented in Table 6.

Table 6. Different income source for the household

Wage Employment	Self-Employment	Non-labor Income
Private Formal	Own business enterprise	Remittances and transfers
Government or parastatal		Income from property ownership
Farming (not own farm)		
Wage Employment		

Unit of analysis

The unit of analysis for the income diversification index depends on the choice of the index. In this case we focus on the inverse of the Herfindhal's index of concentration (2006). This measure is unitless and takes on a minimum value of 1 for the least diversified source to an upper value that is dependent on the number of income sources and shares. The higher the diversity score, the more income diversified the household is. The equation or algorithm used for estimating this score is: $S_j = \left(\frac{Y_j}{Y}\right) \text{ where } Y = \sum_{j=1}^{n} Y_j$

$$YDIV = \sum_{j=1}^{n} \left(\frac{1}{(S_j)^2} \right)$$

where Y_j is income from source j, Y is total income from n sources, and S_j is the share of income from source J (for additional details, see Block and Webb, 2001; Ersado, 2006 Minot et al., 2006).

Limitations regarding estimation and interpretation

Computation of this index is data intensive and may complicate measurement where data on income are difficult to enumerate for labor or where data quality issues exist. Ersado (2006) recommends an alternative simple measure of the number of income sources per household. Although not taking into account the share of each income source, it does provide a count of the number of income sources that household has. This may be computed by simply listing the income sources and asking the household to indicate which of those are activities from which the household obtains income. The maximum value of this index will be the number of activities listed.

Returns to Land, Labor and Capital

Description of the indicator

A number of factors, such as land and labor, are used in the production process. It is important to examine the contributions or returns to these factors so that a farmer may assess the benefits of using an additional unit of a factor when there are competing wants.

In this section, we will focus on assessing the computation of returns to labor for a technology, drawing on work from the net income or gross margin calculation. Profitability metrics can be used to assess what economists refer to as returns to factors of production (e.g., labor, capital, and management; see Engle, 2012).

Method of data collection and data needed

Data used to calculate the returns to land labor and capital are similar to those used to calculate net income. It is important to note that in cases of examining factors such as labor, data on the amount of labor used and the wage rate or value of that labor are collected.

Unit of analysis

Net income is calculated as illustrated in Metric 1: Net Income. To calculate the net return to factors of production like labor, we use the following equation:

$$Net\ Returns\ to\ labor = (Net\ income + \ Variable\ Labor\ Costs)/Total\ labor\ time. \quad (2)$$

Net returns to labor provide an indication of how much that factor[4] (labor) earns in that activity. This may assist the farmer to assess the opportunity cost of labor relative to another activity or technology.

Poverty

Description of the indicator

Poverty is estimated as a welfare measure to ascertain the minimum level of income that is adequate to sustain a livelihood. The measure provides an indication of the standard of living of a person or household in a given location using a monetary metric measure. In socioeconomic and agricultural research, this measure is used to categorize households as "poor" or "resource poor" so that households are identified for proper targeting.

Metric 1: Poverty rate (per capita consumption expenditure)

Description of the metric

Poverty rates have been used for over the past half century to provide a monetary measure of household income relative to meeting the household's basic needs. An international poverty line is set by the World Bank and is used to calculate the number of persons that fall below or above this measure. The World Bank has defined extreme poverty as living below $1.25 (PPP) per day. Rather than income, poverty rate is mainly calculated in developing countries using consumption expenditure to obtain a wealth status of a given (Deaton, 1997). To obtain a per capita measure of consumption expenditure, the value of consumption expenditure by the household on food and non-food items is then divided by the number of household members. Poverty indicators, such as 1) head count ratio (percentage of people below the poverty line), 2) poverty gap (estimates of the 'depth' of poverty), and 3) measure of income inequality among the poor ("gini index"; see Foster et al., 1984) are then estimated.

Method of data collection and data needed

The most popular approach for measuring poverty at the household level is the use of consumption and expenditure surveys. To calculate per capita expenditure through consumption expenditure, the following data are required;

1) Food consumed at home (value of food: price and quantity of each item in local units and currency);
2) Food purchases (quantities and prices of food purchased for consumption, for example 2 kg of meat at a price of 3000 shilling per kg or purchased meat at a value of 6000 shillings);
3) In-kind food consumption (value of in-kind food, e.g., maize flour provided to the household by a neighbor);
4) Household demographics (size, age, gender);
5) Adult equivalence scales;
6) Non-food purchases (e.g., wood fuel, petroleum);
7) Consumer price index (CPI) for the respective country in specific currency (current year local CPI and local CPI for 2005);
8) United States dollar purchasing power parity for 2005; and
9) International poverty line.

Once the data on food consumed at the home, food purchases, and in-kind food consumption have been obtained (see NBS, 2014; section J: consumption of food over past one week), determine the value of each source of food procurement. Sum up the value of each item for each household. The output is the household number in the column and the values in local currency of the 1) value of food consumed at home, 2) value of food purchases, and 3) value of the in-kind food consumption. These values are normally reported using a 7-day recall period to obtain the total household expenditure on food items, sum up the three values for each household (food consumed from own production, household purchases, in-kind food values), and then multiply this value by 52 for an annualized value. For each household, we have a dataset for "household food expenditure (by value)".

The next step is to calculate the value of non-food purchases. This is termed "Non-Food Expenditures" in the LSMS household survey of Tanzania 2012-2013 (NBS, 2014) as an example of the items it constitutes.

First, determine the monetary value for each non-food item purchased., . These are mainly listed in terms of their recall period. This will be needed to calculate the total annual expenditure for that item. For example, if a household spent $100 a month on electricity and $10 a week on public transportation, a different annualization value will be used.

From this step, once the value of all non-food purchases has been calculated and stored, we annualize the value of each item by multiplying that value by the recall period (see Table 7). Sum all the values for the non-food purchases for each household to obtain the "household non-food expenditures (by value)".

Table7. Annualization values

Recall period	**Annualization**
Daily	Multiply by 365.24
Weekly	Multiply by 52
Monthly	Multiply by 12
30 days	Divide by 30, then multiply by 365.24
12 months	None; or multiply by 1

Next, calculate the household's adult equivalence from the demographic data on the household members(e.g., age and gender). Different countries have different scales and Table 8 shows an example of one from Tanzania (Collier et al., 1986). Each household member is first grouped into these categories by age and gender and then multiplied by the adult equivalence weight. These weights are summed up across the household to obtain the household's total adult equivalence.

Table 8. Adult equivalence scale conversion for Tanzania

Age Group [x, x]	Male	Female
0-2	0.4	0.4
3-4	0.48	0.48
5-6	0.56	0.56
7-8	0.64	0.64
9-10	0.76	0.76
11-12	0.8	0.88
13-14	1	1
15-18	1.2	1
19-59	1	0.88
60+	0.88	0.72
Source: Collier et al., 1986		

At this stage, you will have calculated the:

1) Households adult equivalence,
2) Household food expenditures,
3) Household non-food expenditures, and
4) Household size.

From 1 and 2 above, sum up in local currency the non-food expenditures and household food expenditures to obtain the household total expenditures. To obtain the per capita household expenditure, divide the household total expenditure by the household size. If you want to obtain the per capita expenditure per adult equivalence, divide the total household expenditure by the household adult equivalence.

At this point, we need to convert this value to 2005 USD and compare it with the 1.25 poverty line (please note that these lines may change but the UN and World Bank will publish these data). To make this conversion, you need the purchasing power parity exchange rate for that country (see http://iresearch.worldbank.org/PovcalNet/PovCalculator.aspx) and the CPI for that country (e.g., Tanzania) for the year of the survey and for 2005.

To deflate the value of that per capita household expenditure to 2005 local currency (LC) we compute

$$Income\ in\ 2005\ LC = per\ capita\ household\ expenditure * \left(\frac{CPI\ in\ 2005}{CPI\ in\ year\ of\ survey}\right). \qquad (3)$$

Then to convert this to USD 2005 using the purchasing power parity conversion

$$Income\ in\ 2005\ USD = \frac{Income\ in\ 2005\ LC}{PPP\ 2005\ in\ local}. \qquad (4)$$

Equation 4 converts that per capita expenditure to USD for comparison with the 1.25 USD per capita line.

Unit of analysis

The unit of analysis is the per capita consumption. The literature provides algorithms for analyzing data to estimate the number of households below the poverty (Deaton and Zadia, 2002; Deaton 1997; Foster et al. 1984) and have also been applied by organizations such as World Bank in their current analysis of LSMS data sets. To estimate the head count ratio, or percentage of household below the poverty line, the FGT(Foster et al., 1984) algorithm is used:

$$FGT_0 = \frac{H}{N}$$

where H is the number of households below the poverty line (1.25 USD) and N is the total number of households in that sample.

Limitations regarding estimation and interpretation

Poverty rate estimation may be expensive since it requires data collection across various food and non-food items. Measures such as the asset-based wealth index may be less expensive proxy measures for wellbeing. The Progress Out of Poverty measure (Desiere et al., 2015) has been proposed but it suffers from low variability over time and might work best for one-time ‘quick and dirty’ measures. It is similar to the asset-based measure in terms of data collected.

Metric 2: Asset index

Description of the metric

An asset index is a proxy measure for the economic wellbeing of a household (Sahn and Stifel, 2003). In cases where consumption and expenditure data are not available, an asset index is often used (Filmer and Pritchett 2001; Carter and Barrett 2006). This approach is argued to be a better measure than consumption or income since it is more stable over time (Carter and Barrett, 2006; Michelson et al., 2013). It provides a relative measure of poverty for each household and may be used to complement consumption-based poverty measures.

Method of data collection and data needed

Data to operationalize this indicator are obtained by survey. Surveys such as the LSMS and Demographic Health Survey (DHS) contain sections that collect data used to compute this metric. These data include (but are not limited to): ownership of productive assets (bicycle, motorcycle, TV, vehicle, mobile phone); roofing, wall, and floor materials used on the main dwelling; and type of sanitation facilities (for a comprehensive list, see Vyas and Kumaranayake, 2006; Rutstein, 2008). These assets are normally catalogued in the household survey questions about asset ownership--a list of assets is provided for the household to indicate which and how many of them they own . The questions asked are :

1) How many [asset] does your household own?
2) What is the age of the [asset]?
3) At what price did you buy [asset]?
4) If you wanted to sell [asset] today, how much would you receive?

Unit of analysis

Once the data on the items have been collected then proceed to calculate the asset index.

First, tabulate the list of assets by frequency of ownership by households in the sample. If the asset is owned by more than 95% of the households or less than 2% of the households, it is advised to remove them from the list since they will exhibit little variation. The percentage of ownership is a value judgement and may depend on the sample size.

Secondly, run the principle component analysis on the list of assets in order to reduce the dimensionality into a single asset score. It is recommended to use the first principle component that explains the most variance in the data. Using multiple components may add additional complexity in explaining the asset factors (Mckenzie 2005; Vyas and Kumaranayake 2006; Filmer and Prichett, 2001).

The factor scores from the first component are used as weight for each asset in order to construct an asset index for each household. The higher the household asset index score the higher the household's relative economic status in that area (village or sample). Household asset indices are normally stratified into wealth quintiles (Rutstein, 2008), deciles or terciles depending on the study.

The DHS also provides step-by-step information on constructing an asset index if another source is needed (http://www.dhsprogram.com/topics/wealth-index/Wealth-Index-Construction.cfm; http://www.dhsprogram.com/programming/wealth%20index/Steps_to_constructing_the_new_DHS_Wealth_Index.pdf; and Rutstein et al., 2004)

Limitations regarding estimation and interpretation

The asset index is used as a relative measure of poverty among households and does not provide an absolute measure of poverty within a community or across years. The index can help to measure the relative evolution of the household's asset wealth in the community over time compared to other households but cannot assess if one group is poorer or richer in an absolute sense.

Metric 3: Wealth categorization

Description of the metric

Wealth is a categorization used in most studies and societies to assess the relative socioeconomic status of an individual or household in a given setting, context and/or community. Wealth ranking is a participatory exercise where key informants rank households in the community according to their evaluation of each household's resources. The ranking provided from this exercise is like a weighted average of the household's resources and it is important to note that higher weights are implicitly given to resources considered socially more important by the key informant (Kebede, 2009). For this reason, data from wealth ranking may not align with a ranking of households based on a survey of assets. Wealth ranking can thus provide important insights into the social values of community members.

Wealth rankings are often carried out actually as wealth categorizations using pre-defined wealth categories, such as wealth terciles (3 categories) that might be defined as "poor", "intermediate", and "rich". The specific criteria for falling into each category can be developed and made explicit by the key informants. Wealth is a relative category, therefore, research scientists should be aware that members in the communities possess knowledge of the wealth positions of community members.

Analyzing how wealth interacts with technologies is important because the adoption and adaptation of an intervention may differ by the endowment in resources or productive assets. Knowledge of participant's wealth categories may provide information to the scientist to guide recommendations of how the intervention may match the differing resources in the community.

Using this wealth ranking approach may be an important alternative and complement to other methods, such as conventional surveys that estimate poverty rates and asset-based wealth indices, and may be cheaper to administer.

Method of data collection and data needed

Wealth ranking could be carried out with a key informant using a card-sorting technique to categorize households by wealth. A first step in collecting these data is to identify the study area (e.g., village or ward). Community selection is an important part of the process and researchers should aim to carry out the exercise in communities that represent the diversity of conditions (agro-ecological as well as socioeconomic) in their focus area. Because the exercise requires a key informant to be familiar with all of the households, an upper limit of 100 households is suggested (Grandin, 1988).

A list of households in the village should be compiled taking into consideration how the "household" is defined. The most important unit to rank (e.g., nuclear household, extended household) will depend on how access to resources is organized (Schoonmaker Freudenberger,2008). For example, in many parts of Mali the nuclear family is not typically considered a household because land is allocated and labor coordinated by the extended family. The definition of "household" should be discussed with the informants who will assist with the ranking. The names for each household should be written on a small card and a unique household number should be given to each card.

Grandin (1988) states that "obtaining a complete list of household heads is the most difficult aspect of wealth ranking" (p.13). A few residents of the community should be asked to "mentally 'walk' through the area giving names of households living in each place" (p.14). Special effort may be needed in transient communities (such as pastoralists). Even in sedentary communities it is recommended to inquire about other households living in the community but who do not consider it their "home" (Grandin, 1988).

The informants should be long-standing community members familiar with the individuals in the area and known for being honest. It is recommended male and female informants be selected representing a cross section of the community. Grandin (1988) does not recommend using community leaders or extension agents, but rather "ordinary farmers".

Interviewing informants may be done as a group or individually. Grandin (1998) recommends at least three independent rankings be completed (i.e., three individuals or three separate groups) to reduce bias. Names of each household should be written on a card and several key informants (at least three) should place the cards in piles according to the wealth of each household (Grandin, 1988). The cards should be shuffled between each interview. Bellon (2001) proposes using several informants to work in a group aiming at consensus in how they categorize the households. This simplifies the analysis because it reduces the likelihood of divergent responses for any given household. However, such a group process could easily be biased by an influential or charismatic informant. Grandin points out that it generally takes longer for groups to complete the ranking because of the time it takes to discuss categorizing each household.

There are two major components of the interview – the actual sorting of cards and a discussion about the characteristics of the households in each category. Practitioners vary in which component is carried out first. Grandin (1988) recommends allowing the informant to decide on the number of piles, even splitting or combining piles as they sort the cards, and then at the end discussing the characteristics of each group. Other authors recommend asking the key informants to decide on the categories and characteristics of each category before sorting (Bellon 2001; Heemskerk et al., 2003).

Sorting first allows for more of a natural grouping that emerges from the specific households. The subsequent discussion is based on the informants' specific thinking as they categorize the people they know. Establishing categories first, in contrast, draws on the informants' theoretical framework of wealth categories, which may be difficult if it is something they have not explicitly considered before.

During the sorting, Grandin (1988) recommends that if an informant seems hesitant about a household it is better to have them leave it as "unknown" rather than forcing them to categorize it, perhaps inappropriately. She also recommends reviewing each pile once all the households are sorted to double check the grouping and allow the respondent to adjust the final placement of any household.

Informants typically use four or five piles, though some may form as many as nine groups. If an informant puts most households into one pile, encourage them to divide it into two or more groups. Grandin (1988) recommends no pile having more than 40% of the households. Bellon (2001) suggests using three simple categories – poor, intermediate, and wealthy.

When the number of piles is not predefined then after the ranking is completed there is the opportunity to inquire about the informants' criteria for grouping the households into those piles. Researchers should ask the informant what makes households in one group different from households in the other groups. Grandin (1988) recommends starting with the richest group and asking how they are similar and what they have that makes them rich. After having the informant describe the characteristics of each wealth category, the researcher has an opportunity to inquire about their specific research interests – for example, what challenges each group faces in their crop or livestock production (Grandin, 1988).

When establishing categories before ranking the cards, the interviewer should ask the informants to define what is meant by "wealth" in their setting or village. Heemskerk et al. (2003) recommend asking respondents for indicators of wealth or standard of living and then discussing the systematic causes of variability among households. After this is ascertained, the interviewer should discuss with the informant the characteristic of a "rich" household and then those of a "poor" household. The interviewer should then note the characteristics of each group (i.e., the "rich" and "poor" households) on a legible chart, and discuss it with the informants to see if they are in agreement. The key informants should discuss and describe the various categories in order to achieve consensus about the number of categories and their designation (e.g., 'low-resource' and 'high-resource' households).

The information from the card sorting and the related discussion should be recorded. A sample recording sheet adapted from Grandin (1988) is in Box 4. For each group, the household numbers should be listed as well as notes about the characteristics of that group.

Wealth ranking, although indicated as an "alternative" method, may be complementary to data collected via other sources, like household surveys. Kebede (2009) carried out both methods with the same households in four East African countries and found wealth ranking to be important for understanding the social values of resources. Scoones (1995) carried out both methods in Zimbabwe and found the ranking to be correlated with livestock ownership and crop production and an accurate estimator of relative wealth. Furthermore, Scoones notes that the qualitative information obtained during the ranking "reveal[s] details of the historically, socially and economically constructed understandings of wealth and well-being of different actors" (p.67).

Box 4. Wealth Ranking Data Collection Sheet
Informant: ___________________________ Date: _____________________________
Age: _________________________ Community: ___________________________________
Group 1 (Richest):____
____
____
Group 2: ____
____
____
Group 3: ____
____
____
Group 4: ____
____
____
Group 5: ____
____
____
Group 6: ____
____
____
Group 7: ____
____
____
Unknown households: ____
Comments:

Unit of analysis

Grandin (1988) recommends the following scoring method to develop average scores for each household:

1) Give each household a score for each informant based on the pile in which the household is assigned. This can be accomplished by dividing the pile number by the total number of piles:

$$WS = (P/N) \times 100$$

   where WS is the household wealth score, P is the pile number and N is the total number of piles. For example, if the richest group is pile 1 and there are 5 piles, then all the individuals in that group are given the score 20 (1/5 = 0.2; 0.2 X 100 = 20). Individuals in the poorest group will have a score of 100.
2) Put these scores into a table with a row for each household and a column for each informant. Calculate the average score for each household. Note any extreme cases (where one informant ranks a household as wealthy and another informant ranks it as extremely poor) and follow up with the informants to reconcile the difference.
3) Sort the households by their average score.

4) Break them into at least three wealth groups as appropriate. An equal interval is calculated by subtracting the lowest average household score from the highest average household score and dividing by the average number of piles.
5) Calculate the percentage of households in the community in each group. If the research aims to track changes in poverty rates over time, then informants should be asked which categories are considered poor and which are not considered poor.

Limitations regarding estimation and interpretation

The primary limitation of the wealth ranking is that it provides a relative assessment of wealth. One complication is that it is difficult to see how a household has changed over time because their relative position in the community could change even if their physical assets do not (for example, if the wealthiest community members moved away during that time). It also makes wealth ranking comparisons to other communities difficult since the categorization of a "rich" household in one community may differ from that in a different location (Barahona and Levy, 2007). However, Bellon (2001) has found that the characteristics informants use to classify village households may provide a rough idea for comparison with households across villages.

Input Use Intensity

Description of the indicator

Input use intensity measures the amount of a given input used per unit of area (e.g., kg nitrogen input/ha; liter of irrigation water/ha). Input intensity provides a measure of assessing two important issues: 1) whether a given input is used and 2) amount used per unit area. This indicator may also provide a sense of the community's input usage and assess the need to supply more (in cases where low levels are used) or in cases of overuse, explore options to advise farmers about how to reduce the amounts used or use the resource more efficiently. This might include information on diminishing marginal returns and yield plateaus with input use. It is important to note that input use efficiency tends to be highest at very low production levels. The indicator is primarily useful for comparing efficiency of systems with relatively similar levels of production. Input intensity data are mainly measured at the field scale in the agricultural survey (see LSMS-ISA survey).

Method of data collection and data needed

Household agricultural surveys on management practices and output on given areas are a commonly used data collection method. This information is measured at the field/plot level. Farmers may be asked to keep journals or diary entries for specific plots to assess input use. Data collected for this measure are:

- Input used in production. This will be, for example, chemical fertilizer in kilograms, organic fertilizer in kilograms, labor in hours per season note that even though the example has focused on chemical fertilizer use, the concept can apply to other inputs);

2) Area (in hectares) used to produce crop of livestock ;
3) Production in kilograms or tones; and
4) Conversion table for nutrient concentration of common fertilizer blend or mix.

Unit of analysis

The unit of analysis is kilogram (kg) of input per hectare (ha). Use of a simple equation that divides amount of input (in kg) by the area input (ha) is applied for this assessment (Shriar et al., 2002).

Limitations regarding estimation and interpretation

This metric is important in cases where there is a need to monitor or increase input use to reach a targeted level (like the Abuja Declaration of 2006 that proposed to increase fertilizer use from 8kg/ha to 50 kg/ha). Measurement of this indicator in isolation may not provide a holistic outlook on the impact of fertilizer use on productivity unless the major goal is to observe increases in input intensity.

Labor Requirement

Description of the indicator

Labor is a major factor of production in developing agrarian communities with low rates of mechanizations. Examining the labor requirements for growing a given crop and the impact of new technology on labor demand and supply is essential to ensure there is enough labor availability in the season for agricultural production.

Metric 1: Labor requirement

Description of the metric

Quantifying the time spent for particular activities provides an objective comparison of labor requirements when evaluating agricultural technologies or management practices. Additional costs of labor, a critical resource to the household, is then evaluated to ensure that net benefits with the new innovation are realized by the household.

Measurement Method 1: Farmer recall of labor used

Method of data collection and data needed

Data are collected by survey on total number of hours or days that the farmer allocated to produce a given crop. The data are collected for major production activities. In the case of crop production, the ISA-LSMS survey breaks down the activities into the following categories: 1) land preparation and planting, 2) weeding, 3) ridging, fertilizing, and other non-harvest activities, and 4) harvesting (World Bank, 2016). The survey details data for family labor allocated to activities. For hired labor, however, hours worked plus the amount paid for the labor are computed. The payment is made either in cash or as in-kind income or goods. For the latter, an estimation of the value of in-kind goods is important because this may be used to estimate a wage rate. Typical questions may be:

1) During the past [TIME PERIOD] how many days did [household member] spend working on the following activity on this plot?
2) Please indicate the typical number of hours in a day that a person spends working on this activity on this plot?
3) For hired labor: During the past [TIME PERIOD] how many days did your household have hired labor on this plot?
4) How much was paid for this hired labor?

These questions for data enumeration using a survey are available in most World Bank LSMS surveys and can provide a guiding point on how to catalog these data (NBS, 2014; World Bank, 2017).

Unit of analysis

The unit of analysis is the labor hours per hectare allocated to producing a given crop or labor hour per head (cattle). The main issue that comes up in analyses of these data at the household level is ensuring that all data on all household members and hired labor who worked on the field are included. In cases where different individuals worked on a farm, an adjustment of their work contribution to 'woman or man days' should be incorporated. For example, in the LSMS, questions are asked on the typical number of hours a given person works per activity.

For example, if a household head indicated t he worked three days weeding a field; one should inquire about how many hours are in a typical day of weeding. If the household head answers three hours, then this can be used to calculate the number of hours for weeding (nine hours). This should be done for different categories of household individuals, such as children who are involved in farm labor.

Limitations regarding estimation and interpretation

An issue with this metric is the recall period in which the farmer has to provide accurate data on labor use. In most cases, standard surveys collect data at the end of the season or agricultural year, and the information may depend on the correct recollection of information by the informant. If possible, use of journals or cell phone surveys during the season or after a given activity should be used to collect more accurate information (Arthi et al., 2016). There is a major limitation in assessing labor requirements on fields that are intercropped or mixed cropped. In such cases, it might be difficult to obtain a good estimate of the labor allocated to the crop that is being assessed. Use of land proportion to adjust for this may be recommended.

Measurement Method 2: Direct observation

Method of data collection and data needed

Direct measurement method entails an enumerator directly observing the farmer performing a given agricultural activity and entering the data on the duration or time taken and amount of labor (man, machine) used to complete the activity.

Unit of analysis

The unit of analysis is number of hours per hectare for that activity.

Limitations regarding estimation and interpretation

One of the limitations of the direct measurement is the "Hawthorne effect" where the farmer may adjust his work rate when he is aware of being monitored. This may increase the work rate and bias the data upwards. Direct observation would be very costly for a large sample of farmers where activities are performed across a long season.

Metric 2: Participatory evaluation of labor requirement

Scientists may use participatory ratings of labor requirements to collect data on new technology. Farmers participating in the new technology may be asked to provide information on the labor demands and supply. Labor calendars can also provide information on the timing of requirements for labor relative to its supply.

Method of data collection and data needed

Scientists would need to focus on questions of labor relative to the 'local' or current technology. Where the proposed or new technology is used, the questions would have to mainly focus on the amount of labor required to produce the output with the new technology and whether the farmers would find enough labor for the technology for that season. One method to assess labor requirements of new technology relative to the 'local' technology is to ask farmers about their perceptions:

1) Compared to the local technology, how would you rate the labor requirements of the new technology?

2) Does the new technology require more labor?

3) If so, what activity is the most labor demanding for the new technology?

4) Given the labor requirements, how would you rate the local technology versus the new technology?

The rating of the technology may be done using the rating of labor requirements at: "high", "intermediate", or "low". Please note that ratings such as "very good", "intermediate", and "poor" may not be appropriate for this metric so attention should be made towards correct specifications of rating terminology. The use of card charts indicating the activities where labor is demanded for the new technology is critical. Using these activity charts, participants may indicate labor requirements of the technologies. It is recommended that the farmers are familiar with the proposed technology. This may be done by educating them on the attributes of the technology beforehand, or choosing farmers who have experimented or observed the technology being used in the field.

Unit of analysis

The unit of analysis is the average rating. A high rating in this case will mean a high labor requirement and a low rating will mean a low labor requirement. Examining variation between groups of farmers by gender, wealth rankings, and other characteristics may be completed depending on the method of data collection and size of the sample. For example, if a focus group is used to obtain data on labor requirements without recording results for each individual and their demographics (gender), then disaggregation may be difficult. At the same time, if research can randomly sample farmers to obtain a representative population of farmers to participate in the participatory technology evaluation of labor, then statistical tests may be performed to assess if significant differences may exist across groups.

Limitations regarding estimation and interpretation

A limitation to this method may be the sample size and the particular questions the researcher is trying to assess. If the number of farmers is very small, it is very difficult to generalize the findings from this analysis. This method requires, like most data collection approaches, proper training and understanding on the design of the survey questions, use of visualizations, and ensuring the participants understand on how to assess and answer the questions using aids (such as card, stones, grains). Failure to do this might lead to spurious results (CIMMYT, 1988; Bellon, 2001).

Shortcomings regarding estimation and interpretation of the indicator

Labor requirements should be assessed across the growing season. It is important to note that labor demands may vary over the season and by activity depending on demand and supply of labor at that given time. For example, during the planting season, households with large farms may demand more labor and households with smaller farms may have excess labor to supply to the market at that time, consequently covering the excess demand. At the same time, demand for labor across different crops or livestock may differ. Ensuring that peak labor demand activities during the season are noted may be important for farmers to satisfy seasonal variation in demand and supply of labor.

Market Participation

Description of the indicator

Market participation examines whether a farmer sells agricultural commodities or buys inputs from the market. For this metric, we focus on the sale of output to the market. Our interest is in whether the farmers using a given technology are selling the output to the market and, if so, how much they are selling.

Metric 1: Percentage of production sold

Description of the metric

This metric examines the extent of farmer participation in selling goods to market versus keeping them for home consumption. This may provide an indication as to the farmer's choice of output use or raise questions as to why the households are unable to sell surplus production to the market. Depending on whether the good is meant to be a food crop, cash crop or both, this indicator is useful to also examine which type of households are market participants and which are not.

Method of data collection and data needed

Household-level data are collected by survey to compute this metric. The basic data gathered are centered on the amount farmers produce and sell. In the agricultural survey, questions are enumerated on the amount of output that a farmer produced on a given crop or livestock product. This is then followed up with questions to indicate how much or what proportion was sold to the market.

1) Did you harvest any [crop] during the past season?
2) How much of [crop] was harvested?
3) Did you sell any of the [crop] produced during the past season?
4) What was the quantity sold?

A complete set of questions can be found in sections 4A and 5A of the Tanzania LSMS-ISA survey (NBS, 2014) and other standard agricultural World Bank surveys.

Unit of analysis

The metric for this indicator is obtained by dividing the amount that was produced by the household by the amount that was sold for given output. This provides a percentage of amount sold. Therefore, households with a higher percentage participated more in the market for that given good than the others.

Limitations regarding estimation and interpretation

Market participation may be a good basic indicator for examining levels of market participation. It may tell us if households are able to sell surplus goods, which may be a crude way to assume they may recoup their investment in technology. But market participation should be interpreted with caution. High market participation does not mean it will lead to profitability of a technology or that an increase in participation may be an indication of better market access. These issues should be further investigated by the researcher in the context of the project.

Market Orientation

Description of the indicator

Market orientation is defined as the production with an intention to sell to the market. This indicator is used to distinguish between the market participation nature of a household versus the degree of commercialization. Market orientation levels of cash crops are by definition higher than those of non-traditional cash crops, and therefore are listed in metrics 1 and 2. Cash crop production in terms of area may be a proxy to household overall market orientation. However, if a farmer is strategic in his production of crops with intent to sell, the market orientation index -- a composite measure of sales by activity-- will be able to assess the household's output orientation.

Metric 1: Percentage of land allocated to cash crops

Description of the metric

This metric focuses on cash crops and the land allocated to these crops. It is important for the scientist to obtain information that is context-specific on what are considered to be cash crops (these may be traditional cash crops as defined nationally or regionally). The scientists may also have an option to define the cash crop for their study. A threshold may be set as, for example, for crop X, if y% is sold to the market, then it is considered a cash crop. For instance, Frelat et al. (2016) considered an arbitrary number of 90% of total production sold to classify that crop as a cash crop. These levels may be set in consultation with secondary data on area crop sales.

Method of data collection and data needed

The calculation of this metric requires data on the total cultivated land area and the land that is allocated to the production of cash crops. These data are usually collected in the agricultural survey sections on crops grown by plot and allocated area. In the standard LSMS-ISA surveys (World Bank, 2017) the following questions are asked of the farmer:

1) Was the [crop] planted on this plot?
2) Approximately what percentage of the plot was planted with [crop]?

Prior to these questions, these total household land areas were obtained either by visiting the plot and measuring it using a GPS unit or by asking for the farmer's area estimate.

Unit of analysis

The unit of analysis is the percentage of land allocated to cash crops. First the scientists should sum up the total area that was planted with all crops by the household in that season. Next, the scientist should sum up the area that was planted with only the cash crops. The researcher should note if any plots are intercropped and make adjustments accordingly. In the case of a monocrop on the plot, that calculation is straight forward-- divide the total area cultivated by the total area planted to cash crops.

Limitations regarding estimation and interpretation

The market orientation index works best when all the crops grown by the household are included in the estimation. Using a single crop may affect the interpretation of the output. For example, interventions in SSA have tended to focus on staple crops and, although such crops may be income -generating or cash crops for some households (especially the poorer ones), it may be difficult to examine the level of market orientation for this single crop. It is important to discuss with farmers or hold discussion groups to understand the importance of different crops as cash generators or food crops within a heterogeneous population.

Metric 2: Market orientation index

Description of the metric

The market orientation index is used as a measure to distinguish between households that are producing goods destined for the market. The index that we present below gives more importance to agricultural activities that are more marketable. For example, if household A allocates 70% of their total land to three cash crops and the remaining land to food crops, and household B allocates 70% of their land to food crops and the remaining 30% to cash crops, using the market orientation index, the results will show that household A is more market oriented than household B (assuming that food crops have lower marketability than cash crops).

Method of data collection and data needed

Market orientation indices have examined the proportion of household production sold to the market and the proportion of land allocated to crops (Gebremedhin and Jaleta, 2010; Hichaambwa and Jayne, 2012). The data used generate this index can be obtained via agricultural survey. As indicated in the market participation index, data on area planted for each crop will be used to operationalize this metric. In addition, data on the amount sold for each crop should be collected. In standard agricultural surveys data on the amount of harvest sold is enumerated. These data are enumerated for each plot listed by the household and for each crop on that plot. In the LSMS-ISA survey, the following questions are used to obtain that data (for the area questions please refer to the Market Participation section):

1) What was the quantity harvested?
2) What was the area harvested for [crop]?
3) What fraction of crops remains to be harvested in this period?
4) What was the quantity of [crop] sold?

Question 2 may be used in cases where the survey is performed in the same month as the harvest and some crops still remain in the garden, or for crops like tubers where farmers may harvest a portion and leave the other amount still in the field.

Unit of analysis

The index is unitless and ranges from 0 to 1. The higher the index, the more market oriented is the household. The equation for measuring this index is specified in Gebremedhin and Jaleta (2010):

$$\alpha_k = \frac{\sum_{i=1}^{N} S_{ki}}{\sum_{i=1}^{N} Q_{ki}} \quad ; Q_{ki} \geq S_{ki} \qquad \text{and } 0 \leq \alpha_k \leq 1$$

where α_k is the proportion of crop k that is sold, Q_{ki} is the total amount of crop k that is harvested or produced and S_{ki} is the amount of crop k that is sold to the market. The market orientation index (MO Index) is then computed as:

$$MO\ Index_i = \frac{\sum_{k=1}^{k} \alpha_k L_{ik}}{L_i^T} \qquad L_i^T > 0\ and\ 0 < MO\ Index_i \leq 1$$

where L_i^T is the total agricultural land cultivated or operated by household I and L_{ik} is the land that household i allocates to crop or agricultural activity k.

References:

Amir, P. and Knipscheer, H.C., 1989. Conducting On-Farm Animal Research: Procedures & Economic Analysis.

Arthi, V., Beegle, K., De Weerdt, J. and Palacios-López, A., 2016. Not Your Average Job: Irregular Schedules, Recall Bias, and Farm Labor Measurement in Tanzania. Tech. Rep. 21, Universiteit Antwerpen, Institute of Development Policy and Management.

Avila, M., 2016. Technical Paper 5: Concepts and Methods for Economic Evaluation of Alley Farming. http://www.fao.org/wairdocs/ilri/x5546e/x5546e08.htm

Barahona, C. and Levy, S., 2007. The best of both worlds: Producing national statistics using participatory methods. *World Development*, *35*(2), pp.326-341.

Bellon, M.R., 2001. Participatory research methods for technology evaluation: A manual for scientists working with farmers. CIMMYT.

Bellon, M.R., 2001. Participatory research methods for technology evaluation: A manual for scientists working with farmers. CIMMYT.

Block, S. and Webb, P., 2001. The dynamics of livelihood diversification in post-famine Ethiopia. Food policy, 26(4), pp.333-350.

Carter, M. R., & Barrett, C. B. (2006). The economics of poverty traps and persistent poverty: An asset-based approach. The Journal of Development Studies, 42(2), 178-199.

CIMMYT., 1988. From agronomic data to farmers recommendations. *Academic training* manual CIMMYT. Mexico, DF, pp.31-33. (http://repository.cimmyt.org/xmlui/bitstream/handle/10883/830/13803.pdf)

Collier, P., Radwan, S., Wangwe, S. and Wagner, A., 1986. Labour and poverty in rural Tanzania: Ujamaa and rural development in the United Republic of Tanzania.

Deaton, Angus, and Salman Zaidi. Guidelines for constructing consumption aggregates for welfare analysis. Vol. 135. World Bank Publications, 2002.

Deaton, Angus. The analysis of household surveys: a micro-econometric approach to development policy. World Bank Publications, 1997.

Desiere, S., Vellema, W. and D'Haese, M., 2015. A validity assessment of the Progress out of Poverty Index (PPI)™. Evaluation and program planning, 49, pp.10-18.

Dillon, J.L. and Hardaker, J.B., 1980. Farm management research for small farmer development (Vol. 41). Food & Agriculture Org..

Engle, C.R., 2012. Determining the profitability of an aquaculture business: using income statements and enterprise budgets. Stoneville, MS: Southern Regional Aquaculture Center.

Ersado, L., 2006. Income diversification in Zimbabwe: Welfare implications from urban and rural areas (Vol. 3964). World Bank Publications.

Filmer, D., and Pritchett, L. H. (2001). Estimating wealth effects without expenditure Data—Or tears: An application to educational enrollments in states of india. Demography, 38(1), 115-132.

Foster, J., Greer, J. and Thorbecke, E., 1984. A class of decomposable poverty measures. Econometrica: Journal of the Econometric Society, pp.761-766.

Frelat, R., Lopez-Ridaura, S., Giller, K.E., Herrero, M., Douxchamps, S., Djurfeldt, A.A., Erenstein, O., Henderson, B., Kassie, M., Paul, B.K. and Rigolot, C., 2016. Drivers of household food availability in sub-

Saharan Africa based on big data from small farms. Proceedings of the National Academy of Sciences, 113(2), pp.458-463.

Gebremedhin, B. and M. Jaleta. 2010, September. Commercialization of smallholders: Is market participation enough. In Joint 3rd African Association of Agricultural Economists (AAAE) and 48th Agricultural Economists Association of South Africa (AEASA) Conference, Cape Town, South Africa.

Ghana Statistical Service (GSS) 2016. Ghana - Population and Housing Census 2010 - IPUMS Subset http://microdata.worldbank.org/index.php/catalog/2151/study-description Accessed January, 2017

Grandin, B. Wealth ranking in smallholder communities – a field manual. Intermediate Technology Publications, Russel Press, Nottingham, England

Heemskerk, W.; N. Lema, D. Guindo; C. Schouten, Z. Semgalawe, H. Verkuijl, B. de Steenhuijsen Piters and P. Penninkhoff. 2003 A guide to demand-driven agricultural research: the client-oriented research management approach: rural service delivery for agricultural development. Royal Tropical Institute.

Hichaambwa, M. and Jayne, T.S., 2012. Smallholder commercialization trends as affected by land constraints in Zambia: what are the policy implications?. Indaba Agricultural Policy Research Institute.

Hofstrand, D., 2009. Understanding profitability. Ag Decisions Makers, 2, pp.C3-24.

Kebede, B., 2009. Community wealth ranking and household surveys: an integrative approach. *The Journal of Development Studies*, *45*(10), pp.1731-1746.

McKenzie, D. J. 2005. Measuring inequality with asset indicators. Journal of Population Economics, 18(2), 229-260.

Michelson, H., M. Muñiz, and K. DeRosa. 2013. Measuring Socio-economic Status in the Millennium Villages: The Role of Asset Index Choice. The Journal of Development Studies, 49(7), 917-935.

Minot, N. ed., 2006. Income diversification and poverty in the Northern Uplands of Vietnam (Vol. 145). Intl Food Policy Res Inst.

National Bureau of Statistics (NBS) – Government of Tanzania. 2014. Tanzania National Panel Survey 2012-2013, Third Round. Dar es Salaam, Tanzania. NBS. http://www.nbs.go.tz/tnada/index.php/catalog/29

Rural Solutions, S.A., 2012. A gross margin template for crop and livestock enterprises. Farm Gross Margin and Enterprise Planning Guide. By Sagit, GRCD, Rural Solutions SA.

Rutstein, S. O. 2008. The DHS Wealth Index: Approaches for Rural and Urban Areas. Calverton, Maryland, USA: Macro International - See more at: http://dhsprogram.com/publications/publication-wp60-working-papers.cfm#sthash.dhPElqZu.dpuf

Rutstein, S.O., Johnson, K. and Measure, O.M., 2004. The DHS wealth index. ORC Macro, MEASURE DHS.
http://www.dhsprogram.com/programming/wealth%20index/Steps_to_constructing_the_new_DHS_Wealth_Index.pdf

Sahn, D.E. and Stifel, D., 2003. Exploring alternative measures of welfare in the absence of expenditure data. Review of income and wealth, 49(4), pp.463-489.

Schoonmaker Freudenberger, K2008, RRA-PRA a manual for CRS field workers and partners Volume 1. http://www.crs.org/our-work-overseas/research-publications/rapid-rural-appraisal-and-participatory-rural-appraisal

Scoones, I., 1995. Investigating difference: applications of wealth ranking and household survey approaches among farming households in southern Zimbabwe. *Development and change*, *26*(1), pp.67-88.

Shriar, A.J., 2000. Agricultural intensity and its measurement in frontier regions. Agroforestry Systems 49, 301–318.

Swathi Lekshmi, P.S., Venugopalan, R. and Padmini, K., 2008. Livelihood analysis using wealth ranking tool of PRA. *Indian Research Journal of Extension Education*, *8*(2 & 3), pp.75-77.

Vyas, S., and L. Kumaranayake. 2006. Constructing socio-economic status indices: how to use principal components analysis. Health Policy and Planning, 21(6), 459-468.

World Bank, 2016. Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA). http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,contentMDK:23512006~pagePK:64168445~piPK:64168309~theSitePK:3358997,00.html

World Bank 2017. Living Standard Measurement Study Survey Integrated Surveys on Agriculture (LSMS-ISA).

Environment Domain **Part 1: Biodiversity & water** (The superscript letters ([a,b,c]) after each metric refer to the methods in the right-hand column)

ENVIRONMENT DOMAIN (Part 1: Biodiversity and water)					
Indicator	**Field/plot level metrics**	**Farm level metrics**	**Household level metrics**	**Community/ Landscape + metrics**	**Measurement method**
Vegetative cover	% Vegetative cover by type (tree, shrub, grass, invasive) [a,b] % Burned land [a,b] % Bare land [a,b]	% Vegetative cover by type [a,b] % Burned land [a,b]	N/A	% Vegetative cover by type [c] % Burned land [c] % Bare land [c]	[a] Quadrats, transects, or visual estimate of cover [b] Participatory exercise [c] Remote sensing
Plant biodiversity	Alpha Diversity Index [a,b] # Species or varieties [a,b]	Beta Diversity Index [a,b] # Species/varieties [a,b]	N/A	Gamma Diversity Index [a,b] % Natural habitat [c]	[a] Vegetation sample [b] Transects [c] Satellite images
Pest levels	Pest abundance and severity by type [a,b]				[a] Seasonal transects [b] Traps
Insect biodiversity	# Pollinators [a,b,c] Diversity index [a,b,c] # Beneficial insects [a,b,c]			# Pollinators [a,b,c] Diversity index [a,b,c] # Beneficial insects [a,b,]	[a] Traps [b] Direct observation [c] Seasonal transects
Fuel availability	Fuel biomass (e.g., wood, residues) produced on plot [a,b,c]	Fuel biomass (e.g., wood, residues) produced on farm [a,b,c]	% hh fuel by type (wood, charcoal) [a,b] # months energy security [a,b] Fuel collection time [a] % of hh fuel from farm [a]	% of fuel from off-farm [a,b] Spatial arrangement of fuel sources [b] % of hhs with energy security [a]	[a] Survey [b] Participatory exercise [c] Biomass measurement
Water availability	Irrigation use by crop [b] Soil moisture [a,b,c] % of plants wilting [b,d] Infiltration rate [a,d]	Irrigation use [b] % of fields wilting [b, d]	Water sufficiency [b] Water security index [b] Water security rating [d]	% of irrigated land [b, e] % flow not diverted [f] % hh with sufficient water [b]	[a] Field and lab tests [b] Survey [c] Crop models [d] Participatory exercise [e] Remote sensing [f] Stream sampling
Water quality			Rating of water quality [b]	% Clean water sources [a,b] % Pop. w/clean water [b] Salinity [a] Phosphate /nitrate /pathogenic microbe concentration (mg/L) [a]	[a] Water sampling [b] Household survey

Environment domain part 2: Soil and pollution (Note: The superscript letters ([a,b,c]) after each metric refer to the methods in the right-hand column)

ENVIRONMENT DOMAIN (Part 2: Soil and pollution)					
Indicator	**Field/plot level metrics**	**Farm level metrics**	**Household**	**Community/Landscape + metrics**	**Measurement method**
Erosion	Soil loss (tons/ha/yr) [a,b,c] Rating of erosion [a,d]		N/A	Sediment load (mg/L) [e] Erosion (tons/ha/yr) [b]	[a] Direct measurement [b] Models [c] Survey [d] Participatory exercise [e] Stream sampling
Soil biology	Total carbon (% or Mg/ha) [a,c] Labile or 'active' carbon (POXC) [a] and/or CO_2 mineralization [c] Partial carbon budget [b,c] Earthworms [d]	Relative measures of plot-level metrics across farm fields Total Carbon [e]	N/A	N/A	[a] Soil test [b] Survey [c] Modeling [d] Farmer and plant assessment [e] Remote sensing
Soil chemical quality	Soil pH (acidity) [a] Electrical conductivity [a] Soil nutrient levels [a] Nutrient partial balance [b] Biological nitrogen fixation [a]	Nutrient partial balance [b] Biological nitrogen fixation [a]	N/A	Nutrient partial balance [a,b]	[a] Soil tests [b] Survey and lookup tables
Soil physical quality	Aggregate stability [a] Bulk density [a] Water holding capacity [a] Infiltration rate [a]				[a] Soil tests
GHG emissions	CO_2 equivalent emitted per ha [ab]	CO_2 equivalent emitted per ha [ab]	N/A	CO_2 equivalent emitted per ha [ab]	[a] Lookup tables by activity or input [b] Models
Pesticide use	Active ingredient applied per ha [a]	Active ingredient applied per ha [a]	N/A	Pesticides concentration in water [b]	[a] Agricultural survey [b] Water tests

Environmental Domain

Vegetative cover

Description of the indicator

Assessing vegetation and ground cover provides important information that relates to soil conservation, as a high degree of cover (living or residues) is required to prevent erosional forces, such as wind- and water-driven losses of soils. It is also an indicator of habitat for microorganisms and biodiversity, as plants are the primary producers capturing sunlight for carbohydrate production, which is at the foundation of food webs and agroecosystem sustainability. Soil cover measurements in the field are described as part of the Erosion Indicator in the next section.

Metric 1: Percent vegetative cover

Description of the metric

Vegetative cover is the portion of ground area that is covered by vegetation (canopy cover), which may be in natural landscapes or agricultural areas. It can be from a canopy of a variety of things (e.g., field crops, cover crops, trees).

Measurement method 1: Quadrats

Method of data collection and data needed

These methods are described in Anderson and Ingram (1993) Sections 3.1.2 for herbaceous vegetation and 3.1.1 for tree vegetation. For details on the quadrat method, see the Crop Residue Productivity Indicator, measurement method 1, in the Productivity Domain. Several other environmental indicators (e.g., biodiversity, weeds, invasive species) can also be assessed using quadrats and transects.

Measurement method 2: Braun-Blanquet scale of vegetative cover

Method of data collection and data needed

A simple assessment of land cover is available through the Land Potential Knowledge System and associated Land Cover App (https://www.landpotential.org/), which provide a tutorial and advice on how to gather simple information on types of vegetation, both living (trees, bushes, crops) and dead (residues), and calculates indices that can be compared to other sites.

Rough approximations of the area covered by each species or type of plant (trees, grasses, shrubs) can be quickly carried out using the Braun-Blanquet scale described in the table below (Wikum and Shanholtzer, 1978).

Table 1. Braun-Blanquet scale for percent of cover

Braun-Blanquet Scale	Range of Cover (%)
5	75-100
4	50-75
3	25-50
2	5-25
1	<5, numerous individuals
+	< 5, few individuals (rare)

Unit of analysis

The unit of analysis is an observable area of land. This scale could be used for an entire plot or field if the whole is easily observable (i.e., tree cover on a small plot). In many cases, it will be necessary to sample portions of a field or plot, following the instructions for quadrats.

Limitations regarding estimation and interpretation

Because this scale provides a rough approximation of vegetative cover, it is not sensitive to small changes. It instead provides a rapid assessment of vegetative cover useful for discerning major differences.

Plant Biodiversity

Description of the indicator

Plant biodiversity is an important indicator at various scales. At the community and larger scales, the vegetative structural diversity provides important regulatory functions in agricultural landscapes (Newton et al., 2008). The presence of trees and shrubs as boundary plantings and hedges with multiple plant species provide barriers to runoff (for increased capture of water and soil resources), as well as providing important habitat for beneficial insects that often require a high boundary to field ratio (Andow, 1991). Mosaics of different plant life forms across an area are also associated with other important regulatory services; however, metrics at these larger scales are beyond the scope of this manual.

At the farm and field scale, the plant biodiversity indicator provides information about the crop and non-crop populations. Diversification of crops grown, as well as associated plants, involves choices made by a household at the farm level and at the field level. For a given field, crops can be grown in mixed intercropped systems or as sole crops. Non-crop species are sometimes considered weeds and are possibly removed, but often several species are left in the field along with the crops. These non-crop species are also important and should be noted.

The diversity of a landscape is known as gamma diversity, which has two components: alpha diversity and beta diversity. Alpha diversity is simply the number of species in a field type, i.e., the species richness. Beta diversity can be used to compare the diversity of different types of fields and is a useful measure of the evenness of the diversity across an area.

Method of data collection and data needed

We recommend using transects (2m by 50m) across the farm landscape for characterizing the tree diversity. We recommend two to three transects per hectare. The species and number of individuals (as well as tree height and diameter at breast height = 1.3m) is recorded. The landscape herbaceous diversity (e.g., crops, weeds, invasive species, rare species) can also be assessed at the landscape level by placing $1m^2$ quadrats at the 0m, 25m, and 50m positions along the transect. For more details on the quadrat method, see the Crop Residue Productivity Indicator – method 1.

The quadrats used for estimating vegetative cover can be used for estimating field level (alpha) diversity and farm level (beta) diversity. For a particular type of field, maize fields for instance, place five to ten quadrats per field and record the species present. Take the average for the number of quadrats. The species can also be categorized as e.g., crop, weed, invasive, legume.

A major challenge is finding people with taxonomic identification skills to quantify the species-level diversity. It is possible to compare the diversity of functional groups even when identifying to species level is not feasible.

Simple characterization of agricultural diversity can be done from survey data if the data includes the number of crops per household, or the proportion of households or fields with sole crops vs. intercrops, and different crop types (e.g., legumes, cereals, tubers, cash crops, perennials vs annual crops).

At the landscape level, a simple proxy for plant biodiversity is the percent of natural habitat in the landscape. This could be estimated from satellite images or from transects through the landscape.

Unit of analysis

Diversity of a field, farm, or landscape can be compared using a range of calculations. At the most basic level, the incidence of a group of interest may be valuable (i.e., the percent of all trees that are native species).

Beta diversity is often calculated as the number of species unique to the two fields being compared. The larger the number, the greater the differences between pairs of fields and the greater the beta diversity. For a more thorough analysis of different concepts referred to as "beta diversity" and how to calculate, them see Tuomisto (2010).

At the landscape level, gamma diversity is the number of species found along the total number of transects.

Often a diversity index is calculated for comparison of one value representing a balance of alpha, beta, and gamma diversity. There are two main indices of diversity: the Simpson index (which is more sensitive to changes in richness) and the Shannon index (which is a better measure of dominance in terms of abundance). Choosing between the two depends on the goals and context.

Through government records or project survey data, these metrics can be calculated by recording the total number of crops per field per season and the total crop area cultivated. From this data, the proportion of area under legumes, cereals, and tubers can be determined, as well as the proportion of area under sole crop or intercrop, defined as two or more crops per field, per season (e.g., Shaxson and Tauer, 1992).

Limitations regarding estimation and interpretation

Diversity calculations do not reflect the value or function of specific groups or species, so this indicator may best be used in conjunction with other indicators, such as nutrient partial balance, vegetative cover, or pest levels.

Pest levels

Description of the indicator

Pests are organisms that are considered detrimental to human concerns. In the case of crops or livestock, pests reduce productivity and can even cause mortality. Pests include weeds, animals, insects, fungi, and bacteria.

Method of data collection and data needed

Pest abundance and severity by type (e.g., insect, weed, disease, parasite, invasive species) can be assessed using the quadrat method described in detail in the Crop Residue Productivity Indicator – Metric 1, measurement method 1.

Insect pests are determined by different methods. They can be estimated by net sweeps for mobile insects or through crop scouting for pests like aphids or bollworms that would not be reliably caught in a sweep net. Locally developed integrated pest management (IPM) recommendations can be used to identify threshold values.

Participatory approaches can also be used to obtain farmers' assessments of the presence and severity of pests and the potential risks to crop and animal productivity. For details on carrying out farmer ratings, see the Crop Yield Indicator – Metric 2. Those methods can be adapted to ask farmers to rate the abundance and severity of any type of pest.

Unit of analysis

The species and number of individuals of each species should be recorded and expressed as species of "weeds or invasives" per unit area or by field type (e.g., maize fields), as well as number of individuals of each type of insect per area, as an indication of severity.

Limitations regarding estimation and interpretation

Interpreting the impact of pest abundance requires considering the specific timing in relation to crop production and the animal's stage of growth. Populations of many pests can change quickly, so it is important to compare levels across sites at the same point in time or the same phase of crop or animal growth.

Insect biodiversity

Description of the indicator:

This indicator focuses on number of species of pollinators and other beneficial insects, plus the richness of these species (i.e., evenness and abundance).

Method of data collection and data needed

The various metrics for this indicator (i.e., the number of pollinators, the number of beneficial insects, the diversity index) can all be measured using similar methods that will depend on the context and species of interest.

A common method for measuring insect diversity is to carry out seasonal transects through an area of interest using a sweep net. Insects can be identified to the lowest level possible. Information about transects can be found above.

Insect traps may also be useful for specific types of insects. Shining a light on a white sheet at night is one way to attract flying species that are active at night in order to quantify their diversity.

Direct observation may be useful for quantifying the number and types of pollinators or beneficial insect predators seen in a given amount of time in a known area of observation.

Unit of analysis

The species and number of individuals of each species should be recorded and expressed as species of pollinators or beneficial insects per unit area or by field type (e.g., maize fields).

For details on the diversity index, see the Plant Biodiversity Indicator above.

Limitations regarding estimation and interpretation

Comparisons across sites must be done at similar points in time due to the seasonality of insect abundance.

Fuel (energy) availability

Description of the indicator

Many rural households depend on solid fuels for cooking and heating. These sources include wood, charcoal, crop residues, dung, coal, and other forms of biomass. The availability of these sources of fuel can be on-farm or in the landscape. Collecting fuel from the farm and landscape can be done by sustainable means if the production is equal to or higher than the demand. Agroforestry is often a means of providing fuel. In rural landscapes, fuel demand more often is higher than supply, and therefore the farm and landscape are degraded and biomass (and often biodiversity) is reduced. Keeping track of the demand and supply of fuel at the household level can indicate the trajectory of fuel availability and potential landscape degradation or need for reforestation.

Metric 1: Fuel biomass produced

Description of the metric

There are three main categories of fuel produced on-farm: wood, crop residues, and animal dung. Each of these can be quantified at the plot or farm-level. Crop residues and animal dung should only be measured as fuel biomass if they are actually used for fuel in that location or can be used as an estimate of potential sources of fuel.

Unit of analysis

The amount of fuel from crop residue production can be estimated by taking the dry weight of residues produced (see Productivity Domain – Crop Residue Productivity Indicator) and multiplying it by the amount of energy stored per kg. Look-up tables of kilojoules per kg for common woods, crop residues, and animal manure can be used for this purpose.

Wood can be measured by volume (m^3) or by weight (kg). The weight of the wood is more useful because the energy stored in a unit of volume depends on the density of the wood.

Limitations regarding estimation and interpretation

Rural households may not be able to reliably estimate the amounts of fuel they produce because “harvest” is on-going. It may be necessary to ask them to measure amounts for a week or to observe their collection of fuel to obtain a more precise measure.

It is also important to control for fuel quality (i.e., the energy value) when quantifying the amount produced and making comparisons.

Metric 2: Percent of household fuel by type and source

Description of the metric

More detailed questions are needed to understand shifts in household fuel consumption amounts and the sources of those fuels. Brocard et al. (1998) summarize the literature on wood and charcoal consumption for West Africa. On average, fuelwood consumption per capita is 1.3kg wood per day and 4g charcoal per day (Brocard et al., 1998). Another study points out that consumption depends on household size, with larger households having lower per capita fuel consumption due to shared cooking (Cline-Cole et al., 1990).

Method of data collection and data needed

Many household surveys and censuses ask questions about the solid fuels used by households (e.g., wood, crop residues, dung, charcoal, coal) because of the concern about indoor air pollution from cooking on solid fuels (Bonjour et al. 2013). The Demographic and Health Survey (USAID, 2015) contains three questions related to cooking but does not attempt to quantify how much fuel is used.

Figure 2. DHS questions about cooking fuel use

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP
113	What type of fuel does your household mainly use for cooking?	ELECTRICITY . . . 01 LPG . . . 02 NATURAL GAS . . . 03 BIOGAS . . . 04 KEROSENE . . . 05 COAL, LIGNITE . . . 06 CHARCOAL . . . 07 WOOD . . . 08 STRAW/SHRUBS/GRASS . . . 09 AGRICULTURAL CROP . . . 10 ANIMAL DUNG . . . 11 NO FOOD COOKED IN HOUSEHOLD . . . 95 OTHER ______ 96 (SPECIFY)	NO FOOD COOKED IN HOUSEHOLD → 116
114	Is the cooking usually done in the house, in a separate building, or outdoors?	IN THE HOUSE . . . 1 IN A SEPARATE BUILDING . . . 2 OUTDOORS . . . 3 OTHER ______ 6 (SPECIFY)	2, 3, 6 → 116
115	Do you have a separate room which is used as a kitchen?	YES . . . 1 NO . . . 2	

Source: USAID, 2015, available from http://dhsprogram.com/publications/publication-dhsq7-dhs-questionnaires-and-manuals.cfm

In a study to assess how pigeonpea and fuel-efficient stoves affected deforestation, Orr et al. (2015) asked households to estimate the number of times per month they collected fuelwood, the number of family members who participated in collecting fuelwood (to estimate the number of bundles collected each month), and the number of bundles bought each month. The researchers then estimated average weights for head loads (33kg) and for bundles sold in the local market (9.4kg). They note that their questions may not include fuelwood collected from farmers' own trees or woodlots (Orr et al., 2015).

Another study in Malawi explored how wood collection shifted as availability decreased (Brouwer et al., 1997), pointing out how labor availability influences fuel type collected, with smaller households using closer sources of lower quality. Researchers asked each household the type of fuels they used, how they used each fuel (i.e., cooking, space heating, heating bath water, sales), and how much of each type they had in stock. They also asked about fuel purchase and collection (i.e., place, distance, frequency, time required, amount, household member responsible for each). Headloads were converted to kg by directly measuring the most recent headload size. Data was collected during the dry season, the rainy season, and the harvest season (Brouwer et al. 1997).

In Burkina Faso where fuelwood is transported by cart (*charet*), the volume of wood can be estimated at $3m^3$/charet and then monthly consumption per household can be obtained by asking how many carts were collected in a month (Etongo et al., 2016). These researchers also asked about how much was sold and where the wood was collected.

Metric 3: Months of energy security

Description of the metric

Farmers can be asked how many months they typically have enough fuel. This is likely to be a subjective value that can be difficult to compare across sites. For example, an area where fuel sources are distant but abundant may say they have enough, while a site with nearby sources that are declining may say they do not have enough, even if there are abundant sources in a distant location.

Security can be defined more objectively by specifying an amount of time or a distance needed to collect wood, but this can be problematic (see summary of Brouwer et al. 1997 in Metric 2 above); or it could be defined by the prevalence of specific economizing activities (i.e., shortening cooking time).

Metric 4: Percent of fuel from off-farm

Method of data collection and data needed

The percent of fuel from off-farm can be assessed via survey. Household can be asked to indicate the "share of household income spent on fuel and electricity" (Vera and Langlois, 2007) and the source of fuel. From the different source of fuel listed, the percentage of fuel from off-farm source can be calculated.

Metric 5: Spatial arrangement of fuel sources

Method of data collection and data needed

Data on where fuel is collected can be gathered through a survey or through a participatory mapping exercise.

Metric 6: Percent of households with energy security

Description of the metric

At the community or landscape scale, the percent of households with sufficient fuel provides information on the overall availability of fuel. See the discussion in Metric 3 about defining energy security.

Method of data collection and data needed

Similar methods as used in Metric 3 are suggested for this metric.

Unit of analysis

The community, district, or administrative region is the unit of analysis.

Limitations regarding estimation and interpretation

Because this metric is an aggregation of household level data, it is important to ensure consistent understanding of the question by respondents and consistent meaning behind their responses. See Metric 3 discussion about subjective responses to the concept of "adequate" fuel.

Water availability

Description of the indicator

Water availability is of critical importance for both agricultural use and household consumption. The following metrics can be used to measure some of the most critical aspects of water availability.

Metric 1: Irrigation use by crop

Method of data collection and data needed

Household surveys can be used to ask farmers which plots and which crops are irrigated. This information can also be used to identify which households use any irrigation. Researchers will have to decide on the level of detail that needs to be assessed (e.g., the type of irrigation system, the frequency of watering, the quantity of water used). Seasonality is an important issue for distinguishing between dry-season irrigation and supplemental irrigation during rainy seasons.

Limitations regarding estimation and interpretation

Household surveys may be limited in collecting data on variations in irrigation water use within and across seasons. Modification to fit the project and context should be explored.

Metric 2: Soil moisture

Method of data collection and data needed

Research to monitor soil moisture uses specialized probes to take readings at various depths in the soil profile. The timing of the measurement is especially important and detailed studies may require daily readings throughout the growing season. A more qualitative assessment is also possible through visual or touch examinations of the soil at various depths.

Surveys can also be carried out to ask farmers about the timing of inadequate soil moisture on their fields.

Crop models can also simulate the water availability based on the level of plant growth, the daily climate information, and detailed soil characteristics.

Metric 3: Percent of plants or fields wilting

Method of data collection and data needed

A clearly observable metric of the lack of water for agriculture is the percent of plants or the percent of all land where plants are wilting due to drought. This metric is most comparable for one specific crop, as one variety may differ from another in when it wilts.

Farmers can also rate the frequency or severity of the wilting in their fields through a participatory rating exercise. See Metric 2 of the Crop Yield Indicator for details.

Metric 4: Infiltration rate

Method of data collection and data needed

The time it takes for water to absorb into the soil is the infiltration rate. If more water infiltrates into the soil there is less runoff and potentially more water stored in the soil for use by plants. Soil texture is an important factor in the infiltration rate and in how much water is stored in the soil. Comparisons will therefore need to be made on soils of similar texture and depth.

Metric 5: Water sufficiency

Water sufficiency metric is the measure of provision or availability of adequate amount of clean water. This data can be obtained via survey where farming households are asked to indicate water sufficiency for various activities that will include cooking, drinking, and cleaning.

Metric 6: Water security index

Description of the metric

The water security index is based on the premise that households manage their water resources and services in such a way as to satisfy sanitary and water requirements (ADB, 2016).

Method of data collection and data needed

The three key sub-indices used to compute the overall water security index at the household level are: 1) percent of households with access to piped water supply, 2) percent of households with access to improved sanitation, and 3) hygiene (age-standardized disability-adjusted life years per 100,000 people for the incidence of diarrhea) (ADB, 2016).

Metric 7: Water security rating

In the absence of quantitative data on water availability, farmers may be asked to evaluate the level of water security. This can be done by rating or ranking a given season or year relative to a given baseline or target.

Metric 8: Percent of irrigated land

Method of data collection and data needed

The percent of irrigated land can be estimated using household survey data information. See Metric 1 above for more details.

Metric 9: Percent of flow not diverted

Description of the metric

When water is diverted from streams there can be serious ecological ramifications downstream if minimum flows and levels are not preserved. There are often government regulations guiding how much of a stream can be diverted for agriculture or human use.

Method of data collection and data needed

Measuring the flow of a stream can be estimated by measuring the area of the cross section of the stream and averaging the speed of water flow from one bank to the other. A more precise measurement method is constructing a permanent flow meter in the channel.

Metric 10: Percent of households with sufficient water

The measure of percentage of households with sufficient water is obtained once data on sufficient water availability had been computer (see metrics "Water Sufficiency).

Water quality

Water quality indicator describes the concentration of various chemicals of interest. This may include nutrients, hydrocarbons, industrial chemicals e.t.c.

Description of the indicator

Metric 1: Rating of clean water

In the absence of quantitative data or measurements of water quality, farmers or respondent can be asked to rank or rate a given water source or quantity of water for cleanliness. This may provide information on how the respondents differentiate given water sources or quantities in the absence of measurements.

Metric 2: Percent of the population with clean water

This is a measure of the proportion of population with water sources that are considered clean for either human consumption or for agricultural use. This percentage can be obtained from water tests or from farmer rankings or ratings of water sources.

Metric 3: Salinity

Water salinity is a measure of the amount of salt that is dissolved in a water quality or soil (for soil see electrical conductivity metric). High salinity level above 1000mg/liter of salt make water significantly unpalatable for humans.

Metric 4: Phosphate /nitrate /pathogenic microbe concentration (mg/L)

Phosphate and nitrate are essential for life forms. When these nutrients are applied in excess and flow into water streams, they may affect the water quality. These nutrients may increase the growth of plankton and algae which may reduce the dissolvable oxygen for the fish leading to death or in case of algae create a bad odor, taste, and color in water (EPA, 2005).

Erosion

Description of the indicator

Soil erosion is the loss of soil by transport in water or wind. It one of the major types of soil degradation. The effects of erosion are also quite visible on fields as topsoil loss, in fields and landscapes as rill erosion or formation of gullies, or at the landscape level as sedimentation of soil into water bodies. The impacts of soil erosion affect soil processes including loss of topsoil, reduced soil depth and rooting zone, loss of nutrients, loss soil organic matter, loss of biota in the topsoil, and contamination of water resources with nutrients, agrochemicals, and soil. All combined, these changes may affect the primary productivity of entire ecosystems, depending on the soil type, climate, and soil management practices put into place.

Conversion of soils from natural systems to agriculture removes the vegetative cover of soils; in addition, the level of organic inputs that are returned to the soil is reduced. This results in declines in soil organic matter (SOM) and soil aggregate stability, and increases in bulk densities and compaction. These degradation processes result in reduced water infiltration and increased water runoff and soil erosion. The major event that sets off this series of processes is the removal of the vegetative cover, which exposes the soil surface to rainfall impacts and wind. If the soil is also tilled, the soil structure is further disrupted, exposing more of the soil to rainfall and higher temperatures, and leading to further loss in aggregate stability and SOM, and infiltration rates which translate to more runoff and erosion.

Metric 1: Soil loss

Description of the metric

Soil loss at the field level is mainly due to the actions of water, wind, or activities such as tillage. Top soil normally has a high organic matter and loss of these soils may affect productivity and also have downstream effects such as pollution of adjacent wetlands.

Method of data collection and data needed

Erosion caused at the plot or field scale has traditionally been measured with Wischmeier plots, where water (runoff) and soil (erosion) that runs off of the plot are captured at the bottom of the field. These plots are expensive to install and it is costly to collect and analyze the materials. They can be useful for comparing the effects of different soil management interventions on different soils and slopes, but are usually done on research plots by soil physicists and not recommended otherwise. This methodology can result in overestimates of erosion losses, because the possible redistribution of soil among small plots within a field is not accounted for and there could be no net losses at the field scale.

Instead we recommend approaching this metric by estimating the different factors that can contribute to erosion by water. These factors include: landscape features, steepness and length of slope of the field, vegetative cover and soils management, and rainfall erosivity. These factors have been combined into the Universal Soil Loss Equation (USLE or Modified USLE) (Wischmeier and Smith, 1978; Renard et al., 1991) where erosion is estimated or modeled. If measures of these different erosion factors are difficult and the models not available, the relative erosion of different interventions within a site (or between sites) could be compared through visual observations of these factors and ranked as to the most to least important for the different interventions.

Unit of analysis

Tons of soil lost per hectare per year (t/ha/yr) is the unit of analysis for the soil erosion/loss indicator.

Limitations regarding estimation and interpretation

This method may be costly and may likely overestimate soil erosion.

Metric 2: Rating soil erosion

Description of the metric

When comparing different interventions at the field scale, several of the factors that affect erosion are similar, e.g., the slope and rainfall intensity would be similar for neighboring fields. The factors that differentially affect the soil surface exposure and disturbance, however, are important to consider and are related to the vegetative cover and soil management practices.

Method of data collection and data needed

The concept we use here (but not the method) is loosely based on Ludwig and Tongway (1995). Details for this method were modified for the Vital Signs project and are provided below from the E-Plot Biomass Measurements (Vital Signs, 2014).

Table 2. Data collection sheet for soil cover and erosion estimates

Projected canopy cover and ground cover (three to five measurements from randomly placed 1m^2 quadrats. Except where otherwise noted, field analysis is scored from 1-10 as a broad indicator of the percentage out of 100% (e.g., a score of 2 = 20%)						
Intervention name	Quadrat 1	Quadrat 2	Quadrat 3	Quadrat 4	Quadrat 5	Average
1. Canopy cover score (1-10)						
2. Herbaceous crown cover (1-10)						
3. Soil erosion (0, 1, 2 or 3) 0=none, 1=slight, 2=moderate, 3=severe						
4. Surface condition score (0-10)						
5. Rooted plants						
6. Litter cover						
7. Recently disturbed (tilled) - enter negative number for this						
8. %BARE						
9. Bare & porous soil						
10. Bare but sealed soil - negative number of this						
Total (1 to 8 possible)						

Background information for soil quality indicators

Soil Sampling

Before detailing the various soil quality indicators (e.g., measurement of soil carbon, soil chemical and physical quality attributes), this section addresses the methods to obtain a representative sample from the fields of interest.

Box 4. Procedure for field soil sampling

Soil sampling equipment

1. Soil auger (open end or closed end depending on soil type, closed for sand) or soil probe or trowel
2. Bucket, basin or any open type of container
3. Study plastic bags and tins
4. Labels
5. Permanent marker

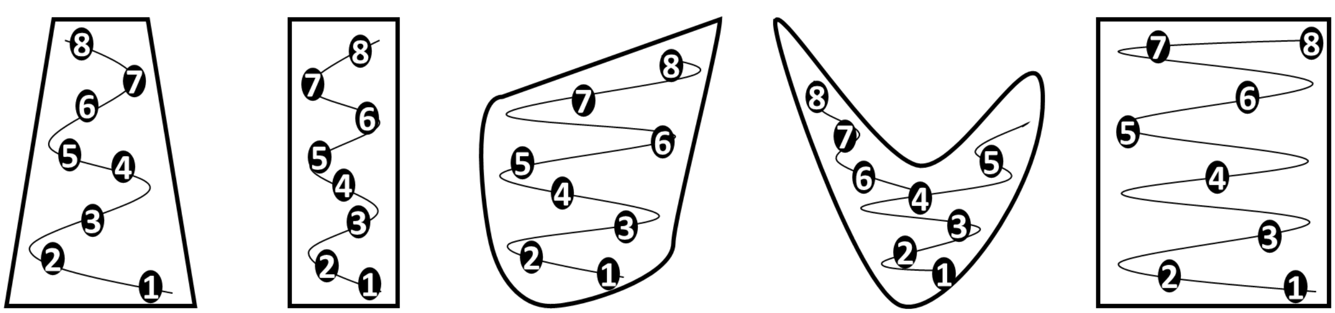
Soil Sampling Steps

5) Familiarize yourself with the plot dimensions
    a. Know where field boundaries are; b. do not sample areas that are unusual (e.g., termite mounds, tree stump areas); c. if the field is large then evaluate to see if more than one soil type is present, based on visual observations of apparent texture and color, then delineate the field based on the main soil types and sample each separately.
6) Collect about five to ten* sub-samples for each field, plot, or soil type area; for a topsoil sample, usually a 0cm-20cm depth is sampled, while subsoil sample is usually from 20cm to 40cm depth. If the plow layer is deeper, then a 25cm or 30cm topsoil sample can be collected; for some research objectives such as investigating soil microbiology, then the top most layer (0-5 cm for example) may be sampled.
   *Choose a set number of sub-samples to be taken (e.g., eight) and keep it consistent throughout the sampling
7) Take sub-samples of soil following a zig-zag path through the plot (see sample diagrams below).
8) Starting at one corner of the plot, first move away from the edge towards the center, and then move through the plot in a random zig-zag manner to collect samples to place in a container and mix well. This is a composite soil sample that represents the plot. Take more samples if the plot is large or if you are measuring inorganic nitrogen, which is often heterogeneously distributed.
9) Samples can be collected with either tool.
    a. Soil auger
        i. Brush aside residues from sampling site (e.g., leaves, plant materials)
        ii. Insert the auger directly into the soil in a **vertical** (straight up and down) position to a depth of 20cm (= 8 inches)
        iii. Carefully remove the auger (avoid any spillage of sample). If soil is dry at sampling time, slightly tilt the auger back to avoid it spilling from tube
        iv. Place the sample in the container and move on to the next sampling site

b. Trowel
    i. Brush aside residues from sampling site (e.g., leaves, plant materials)
    ii. Insert the trowel directly into the soil in a **vertical** (straight up and down) position to a depth of 20cm (= 8 inches)
    iii. Gently push back on the handle and remove the soil (ensuring that you obtain the soil at insertion depth)
    iv. Place the sub-sample in the pail and move on to the next sampling site

10) After all samples are collected, mix up the soil very well and use this soil as the sample
    a. Remove any large stones sticks or roots from the sample
    b. Break up any soil clods with your hand
    c. Mix by hand very well for at least a minute until all the soil is homogenized
    d. Place about one-quarter of the sample (or about 500g) in a bag by using the "pie" method of dividing the soil in the container, by dividing in quarters and collecting one "slice of the pie." It is important to collect a representative quarter or so of the soil; this ensures all layers of soil are collected, as soil will tend to self-sieve by texture. The remaining soil should be returned to field.
11) Label the sample bag with the following information:
    a. Date
    b. Sample ID
    c. Farmer name or number (if number check that the number is correct based on the list)
    d. Location (GPS coordinates)
    e. Treatment
12) Place a small piece of paper with the following information written in pencil into the sample bag:
    a. Date
    b. Sample ID
    c. Farmer name or number
    d. Location
    e. Treatment
13) Securely twist and tie the plastic bag with the sample and store for transport

Figure 3: Recommended sampling zig-zag scheme for eight sub-samples



Source: Snapp et al. Unpublished

Measurements conducted in the field with hand-held monitoring equipment can provide complementary information, or act as a substitute for soil sampling. This is an active area of research and in-situ field soil and plant monitoring is becoming a reality that can be used by field practitioners and scientists (Rossel and Bouma, 2016; Snapp and Morrone, 2008). However, it is important to keep in mind the highly heterogeneous nature of soil and plant properties, both in terms of space and time. This is illustrated by the hand-held devices to assess light reflection, where specific wavelengths are monitored with well-characterized relationships to plant stress and chlorophyll content (and positively correlated with nitrogen content). Careful sampling strategies and dozens of readings per plot are required, using comparable leaf phenology, in order to characterize plant chlorophyll status. This is due to the high variability that occurs at the sub-meter and meter level (Markwell et al., 1995).

Preparing soils for carbon, nutrients, acidity, salinity tests

Once the soil has been sampled it must be dried, cleaned, and sieved before chemical analyses can be carried out. The following boxes provide step-by-step instructions for preparing soil samples for chemical analyses.

Box 5. Soil Moisture Measurement and Drying
If one of the indicators relates to soil moisture, then the soil sample can be analyzed to determine the soil moisture content as follows:

1) Get a soil tin, mark it with a number.
2) Weigh the tin to the nearest 0.1g.
3) Add some of the moist soil to the tin – fill as much as possible.
4) Weigh the tin plus soil.
5) Dry the soils as follows:
    a. Place the opened tin plus the soil with the lid (fit onto the bottom of the tin) into a drying oven at 105°C for 48 hours, or until a constant weight is obtained.
    b. Once soil is dry, close the tin by replacing the lid back on the top of the tin.
6) Weigh the tin plus the oven or air-dried soil, including the lid, to the nearest 0.1g.

Note: The soils in the bags should ideally be oven dried at 105°C. If there is no need to determine soil moisture, then the soils can be air dried. Place the soil on a clean surface (e.g., plastic sheet, shallow bowl), spread the soil out thinly, and put the soil where it will not get wet from rain or contaminated by soil blown by the wind. Mix the soil occasionally to ensure all the soil is dried. Depending on how wet the soil is, the amount of soil, and the climatic conditions, it can take less than a day or several days to dry the soil.
(Excerpted from Vital Signs, 2014b, pp 20)

Box 6. Cleaning and sieving dry soil
In this procedure, the soil sample in the bag will be divided into soil that passes through the 2mm sieve (the fine soil fraction) and the gravel that does not pass through the 2mm sieve (called the coarse fraction). No material should be discarded.

1. Once the soils are air dried, weigh the whole soil sample to the nearest gram. Record the weight.
2. Grind the soils using a wooden rolling pin, gently crushing the sample. While crushing, remove any plant materials (i.e., roots).
3. Remove and save any possible pieces of gravel (making sure they are gravel and not soil aggregates) and place in a separate small plastic bag (coarse fraction).
4. Sieve the soil sample by passing the crushed sample through the 2mm sieve. DO NOT use the sieve as a grinder: do not rub or mash the soil on the sieve, but shake the sieve gently to allow the soil to pass through.
5. Remove and save any gravel that remains on top of the sieve and place it in the plastic bag with the other gravel (coarse fraction).
6. Once the entire sample has been sieved, place the fine soil in a plastic bag (fine fraction).
7. Weigh and record the weight of the fine soil that passed through the 2 mm sieve, record to the nearest gram.
8. Weigh and record the weight of the coarse gravel fraction that did not pass through the 2 mm sieve, record to the nearest gram.

(Excerpted from Vital Signs, 2014b, pp 22)

<u>Soil carbon</u>

<u>Description of the indicator</u>

Soil organic carbon (SOC) is a fraction of the soil organic matter (SOM). SOM integrates many soil properties and can serve as an indicator of the soil's health and level of various soil processes. SOM provides the carbon and energy for soil organisms, and thus also supports the biological functions of soil. It affects the soil's capacity to retain and release nutrients for plant growth by contributing to the cation exchange capacity and through mineralization of organic nitrogen, phosphorus, and sulfur. SOM also affects soil water storage and release and exchange of gases with the atmosphere by influencing soil pore size distribution and bulk density by aggregation of soil particles. Soil organic matter can also reduce the toxicity of certain elements and chemicals through chelation. SOM and SOC content are determined by several factors, including the amount of sand, silt, and clay; the climate; and soil management practices. Clayey soils have higher SOM content than sandy soils and wetter areas have higher SOM content than drier areas. The SOM content is ultimately determined by the balance between the addition of organic inputs to the soil and decomposition. Soil management practices can dramatically affect decomposition rates of SOM and SOC by soil biota.

Metric 1. Total soil carbon

<u>Description of the metric</u>

Soil carbon is a critical indicator of soil quality that is important for soil moisture and nutrient retention and livelihood of soil microbes (Doran and Jones, 1996; Reeves, 1997; McBride et al, 2011).

Method of data collection and data needed

Soil carbon is usually assessed as total SOC through combustion (oxidation) of the soil. This can be done through various methods including burning in a muffle furnace, wet chemistry (Walkley Black), or combustion in CHN analyzers.

Unit of analysis

Total SOC is expressed as units of C per unit of soil (e.g., ug/g, g/kg) or as a percentage; it can also be converted to C soil stocks (t C/ha) when the concentration is multiplied by the bulk density of the soil and the depth of the soil sample.

Limitations regarding estimation and interpretation

Due to the spatial and temporal variability of SOC, it is extremely difficult to detect differences among treatments unless they have been in place for many years – often 10 years or more. The total pool of SOC is quite large relative to the small changes in accrual (or loss) in SOC that can occur over time; when combined with the heterogeneous distribution of SOC, it is very challenging to detect meaningful SOC differences over time. It is also important to consider that soil compactness (bulk density) is problematic to accurately measure, and this poses a major problem for detection of differences in SOC, as the volume of soil sampled is impacted by soil compaction. The analyses are also expensive and require specialized equipment.

We thus do not recommend measuring total soil carbon for experiments or for farmer's fields that have been recently installed. There are some other measurements that provide an indication of more readily decomposed fractions of soil carbon: so-called "Active Carbon," which is described below.

Metric 2: Labile or "Active Carbon"

Description of the metric

Changes in these active fractions of soil carbon appear to estimate the trajectory of total SOC (decreasing or increasing) and can be used to compare different treatments and management practices a year or so after installation.

Method 1: Permanganate oxidizable carbon (POXC)

Method of data collection and data needed

This procedure describes a technique for the determination of oxidizable carbon in soil samples by a dilute solution of permanganate, termed POXC. This procedure is synonymous with the "Active Carbon" method described by Weil et al. (2003) and is adapted with help from J.D. Glover (The Land Institute, Salinas, KA) and M. Barbercheck (Pennsylvania State University, University Park, PA). Active carbon is often considered a labile form of carbon that is more easily decomposed by soil organisms; in contrast to total SOC, active carbon can change quickly with different soil management practices. POXC correlates well with SOC and can be considered a proxy for total SOC. It has recently been proposed that POXC relates to the longer term build up, or storage, of soil carbon (Hurisso et al., 2016).

The procedure for determining POXC in detail is as follows:

I. Instrumentation and Materials:

- Spectrophotometer capable of reading absorbance at 550nm
- Weighing balance capable of accurately weighing ~2.50g of soil to two decimal places (0.01g)
- pH meter calibrated for measurement in the range of ~6.0-8.0pH and NaOH for pH adjustment

- Oscillating (or horizontal) shaker capable of at least 240 oscillations per minute (or 120rpm)
- Magnetic stir plate and stir bars
- Adjustable 10ml pipettor and tips
- Adjustable 100-1000µl pipettor and tips
- (2) Adjustable bottle-top dispensers fitted to a bottle of deionized water and calibrated to deliver 18.0ml and 49.5ml
- 50mL disposable polypropylene centrifuge tubes with caps (Falcon tubes)
- Laboratory glassware for reagent preparation and waste collection
- Labeling supplies such as permanent markers and tape
- Reagent grade Potassium Permanganate ($KMnO_4$; FW=158.03g mol^{-1})
- Reagent grade Calcium Chloride, Dihydrate ($CaCl_2 \cdot 2H_2O$; FW=147.01g mol^{-1})
- Soil standard (sieved and air-dried KBS topsoil for use as a lab reference sample)
- Timer capable of tracking time for two and ten minute intervals

II. Reagent Preparation:

$KMnO_4$ Stock Solution 0.2M (makes 1 liter, 2ml use per soil sample):

1. Weigh 147g of $CaCl_2$ and place in a 1000ml beaker. Add approximately 900ml of deionized water and stir till dissolved. Transfer to a 1000ml volumetric flask or graduated cylinder. Bring to volume with deionized water.
2. Weigh 31.60g of $KMnO_4$ into a 1000ml beaker and add approximately 900ml of the $CaCl_2$ solution. Place on the magnetic stir plate (minimize exposure to light) with gentle heat and stir until dissolved completely. Note: Dissolution may be very slow.
3. Once dissolution is complete, place the probe from a calibrated pH meter into the solution (with continued stirring) and measure the pH. Adjust the pH to 7.2-8.5 by adding 0.1N NaOH, 1 drop at a time (endpoint approaches rapidly). Note that it can be challenging to adjust pH, and note in the record what pH that is achieved. With the $CaCl_2$ solution, adjust volume in a 1000ml volumetric flask or graduated cylinder. Transfer to a brown glass bottle and store in a dark place (stable 3-6 months).

III. Standard preparation:

Four standard concentrations (0.005M, 0.01M, 0.015M and 0.02M) prepared from the $KMnO_4$ stock solution. The standard preparation involves first making a standard stock solution and then diluting each standard stock solution to a final working standard. The following materials will be needed:

- 50mL disposable polypropylene centrifuge tubes
- Adjustable 1.0-10.0ml pipettor and tips
- Adjustable 100-1000µl pipettor and tips
- Adjustable bottle-top dispensers fitted to a bottle of deionized water and calibrated to deliver 49.5ml

Part 1 - Standard Stock Solutions: Use the table below to prepare standard stock solutions. These stock solutions can be prepared in centrifuge tubes or in small brown glass bottles and used for three days (stored in glass and in the dark).

Table 3. Standard stock solutions for POXC

Concentration	**Volume of $KMnO_4$ stock solution**	**Volume of deionized water**
0.005M	0.25ml	9.75ml

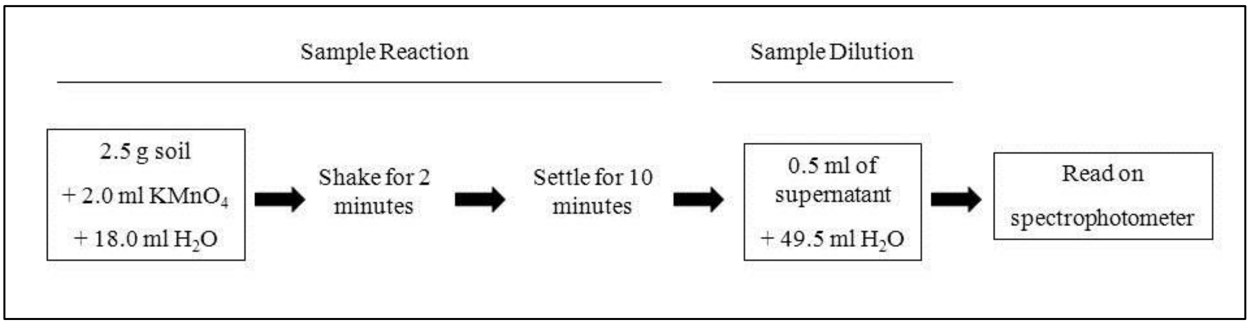
0.01M	0.5ml	9.5ml
0.015M	0.75ml	9.25ml
0.02M	1.0ml	9.0ml

Part 2 - Dilution Step: Dilute each standard stock solution to a working standard by adding 0.5ml of each stock solution to 49.5ml of deionized water in 50ml centrifuge tubes. These tubes contain the working standards and should be prepared fresh daily.

IV. Sample Preparation:

Sample preparation involves a two-part process: a sample reaction and sample dilution, as illustrated below.

Figure 4. Process of testing Active Carbon with POXC



A soil standard and solution standard are prepared in the same manner as the unknown samples. The soil standard serves as a laboratory reference sample. It is recommended to homogenize a large batch of air-dried soil for long-term use. The soil standard allows for a quality control check across POXC analyses performed on different batches, over multiple days, or with different reagents. The solution standard serves as another quality control reference. It is prepared in the same manner as the unknown soil samples, but without the soil. The solution standard will reveal if reagents or labware have been contaminated with oxidizing agents or carbon and thus serves as a true blank.

It is important that the timing of each step be consistent, particularly the shaking and settling times. The permanganate will continue to react with the soil as long as it remains in contact. Hence, working quickly with small batches of 10 samples or less is advised.

The following materials will be needed:

- (2) 50ml disposable polypropylene centrifuge tubes with caps for each sample
- Adjustable 1.0-10.0ml pipettor and tips
- Adjustable 100-1000µl pipettor and tips
- (2) Adjustable bottle-top dispensers fitted to bottles filled with deionized water and calibrated to deliver 18.0ml and 49.5ml
- Labeling supplies such as permanent markers and tape
- Oscillating shaker capable of at least 240 oscillations per minute (or 120rpm) and fitted with a lidded box that will hold at least ten 50ml centrifuge tubes
- Timer capable of tracking time for two- and ten-minute intervals
- Soil standard (pulverized, homogenous soil as lab reference sample)

A. Sample Reaction

1. Label two 50ml centrifuge tubes for each sample. Weigh 2.50g (± 0.05g) of sieved, air-dried soil into one of the centrifuge tubes (may be done in advance). A soil standard should also be prepared. Place the other set of tubes aside.
2. Add 18.0ml of deionized water to each of the centrifuge tubes containing the soil. Using the 1.0-10.0ml pipettor, add 2.0ml of 0.2M $KMnO_4$ stock solution to each tube.
3. Prepare a solution standard by adding 18.0ml of deionized water and 2.0ml of 0.2M $KMnO_4$ stock solution to a tube (no soil) and process in the same manner as the unknown soils.
4. Working quickly, cap tubes tightly and hand-shake each tube vigorously for two seconds to assure soil dispersion within the solution.
5. Place tubes on shaker and shake at 240 oscillations per minute for two minutes.
6. After two minutes, remove samples from shaker and swirl or shake the tube vigorously to ensure that there is no soil clinging to the sides or cap of the tube. At this point, remove caps to avoid further disturbance of soil after settling. Place the samples in a dark area and allow soil to settle for 10 minutes. Settling time is a critical step so a timer is essential.

B. Sample Dilution

1. While samples are settling, add 49.5ml of deionized water to the second set of centrifuge tubes (may be done in advance).
2. Once the ten minute settling period has passed, quickly transfer 0.5ml of supernatant (liquid above the solid, avoiding any particulate matter) to the second tube containing 49.5ml of water. Note: This step should be performed as quickly as possible as the permanganate will continue to react with soil as long as it remains in contact.
3. Cap the second set of tubes and invert to mix. These are the final sample solutions for analysis. They are stable for up to 24 hours if stored in the dark.

C. Reading Samples on Spectrophotometer

1. This method has been shown to perform well on both single cuvette machines and 96-well plate reading spectrophotometers. If available, a 96-well plate reader is recommended to save time (see steps 2-5 below).
2. Clear polystyrene flat-bottom cell culture plates (or equivalent) work well, so more expensive UV-transparent plates are not necessary. Fill each well with 200µl of solution.
3. It is recommended to replicate all standards on a plate, including blanks of deionized water. Running each standard three or more times and taking the average typically yields good results.
4. Determine and record the absorbance (optical density) of standards and unknowns at 550nm using spectrophotometer software.
5. Subtract out average of deionized water blanks from all absorbance values (if not automatically performed by software). The intercept of the standard curve should be very close to zero.

D. Clean-up and Disposal

Leaving the centrifuge tubes capped but on the bench top for a week or more will allow the permanganate to completely react with the soil and lose all purple pigmentation. Liquid can then be safely disposed of down the sink and tubes with soil thrown out or cleaned and reused. The second dilution of samples and standards contains very little $KMnO_4$ and may be safely flushed down the drain with copious amounts of water; however, check with your environmental health and safety department to ensure compliance with your department's procedures.

Unit of analysis and algorithm used for estimation

The amount of carbon oxidized is a function of the quantity of permanganate reduced. Consequently, the higher the POXC values the lower the absorbance (intensity of the color of the solution). Calculating "Mass of POXC for Unknown Soil Samples" can be done using the following equation, after Weil et al. (2003):

$$POXC\ (mg\ kg^{-1}\ soil) = [0.02\ mol/L - (a + b \times Abs)] \times (9000\ mg\ C/mol) \times (0.02\ L\ solution/Wt)$$

Where: 0.02 mol/L = initial solution concentration; *a* = intercept of the standard curve; *b* = slope of the standard curve; Abs = absorbance of unknown; 9000 = mg of carbon oxidized by 1M of MnO_4 changing from $Mn^{7+} \rightarrow Mn^{4+}$; 0.02 L = volume of stock solution reacted; Wt = weight of air-dried soil sample in kg.

Example Calculation:

Construct a standard curve with the values in the following table:

Table 4. Values for creating a standard curve for POXC

Y-axis (Molarity of stock $KMnO_4$ standards)*	0.005	0.01	0.015	0.02
X-axis (Abs values from spectrophotometer)	0.1000	0.1984	0.3034	0.3966

* Note: The standard curve should use the molarity of the stock standards, and not the working standards, since the stock standards represent the actual concentration (0.02 *M* $KMnO_4$) used to react with the soil.

This produces the regression line: y = 0.0502x - 0.00004; $R^2 = 0.999$

Unknown sample absorbance: 0.3087; unknown sample soil weight: 2.48g

$$POXC\ (mg\ kg^{-1}\ soil) = [0.02\ M - (-0.00004 + (0.0502 \times 0.3087)] \times (9000mg\ C/mol) \times (0.02L\ solution/0.00248\ kg) = 329.75mg\ POXC\ kg^{-1}\ soil$$

Limitations of method

This method requires significant equipment and supplies. The procedure is detailed, requiring someone with experience in wet chemistry. Results from this method should only be compared among soils from the same site under different management practices – it is not yet clear that it can be used to compare soils between sites.

Method 2: Mineralizable soil carbon (MINC)

Method of data collection and data needed

Another means of assessing Active Carbon is through measurements of the short-term release of CO_2 from the soil. The amount released and compared among treatments has recently been associated with forms of C in the soil more readily decomposed than those assessed through POXC (Hurisso et al., 2016). It seems to be associated with recent additions of organic matter to the soil and should thus be an appropriate measure (along with POXC) for comparing soils under different management treatments (Culman et al., 2013). Again, these comparisons should be made for treatments or different management situations on a farmer's fields of the same soil type. The CO_2 evolved can be measured several ways. If a gas analyzer is available the method below can be used. One of the most widely used methods is based on an alkali trap that is placed in the container with the soil sample for the incubation period, then removed and titration used to measure the amount of CO_2 trapped. A convenient Solvita method recently has become available that relies on a gel for detection of CO_2 in lab or field-based incubations (Haney et al., 2008).

The challenge with all soil respiration and MINC methods is that for reproducible and comparable results, the soil moisture status needs to be consistent, preferably 50% Water-Filled Pore Space (WFPS). This can be calculated as shown in the Appendix, based on the Haney and Haney (2010) method. Or, a simplified approach that is promoted by Solvita (https://solvita.com/) is to use the equation:

$$50\%\ WFPS\ =\ (volume\ of\ the\ soil\ -\ 40/2.65)\ x\ 50\%.$$

An assessment of how to use Solvita is available from the USDA. See the “Soil Respiration” guide for educators on this website:
https://www.nrcs.usda.gov/wps/portal/nrcs/detailfull/soils/health/assessment/?cid=nrcs142p2_053870.

Box 7. CO_2 evolution from wetting a dried soil (capped IRGA method)

Abstract:

This procedure describes a technique for the determination of CO_2 respired from air-dried soil that has been rewetted. The method is described in detail by Franzluebbers et al. (2000).
The procedure described below uses an infrared gas analyzer. Incubation times can vary from one to three days, or longer. The Snapp lab has found treatment sensitivity and decreased analytical variation with one-day incubations that are closely correlated with results from longer-term (3-,7- and 24-day) mineralization, so for convenience we recommend a one-day incubation.

Materials:

- Ball ® Mason canning jars (half-pint or quart size)
- Lids and rings for canning jars
- Rubber septa for jar lids
- Plastic specimen containers (100ml) or glass beakers (100ml)
- 5ml
- Syringes (1ml and 3ml)
- Syringe needles (25 gauge, 1-inch and 1.5-inch)
- Gas tank – 1% CO_2 (99% N_2) + regulator
- Gas tank – Helium or N_2 + regulator
- Septa for IRGA – SS174 Teflon Faced Septa 11mm (100/pkg) from Supelco Catalog #2-2731 (http://www.sigmaaldrich.com/analytical-chromatography/analytical-chromatography-catalog.html)

Equipment:

- Balance for weighing soil
- LI-820 infrared gas analyzer (IRGA)
- LI-820 software

Prior to set-up:

Drill holes into canning jar lids the size of the septa (or rubber stoppers) and insert septa into holes. Make sure they are airtight. This can be accomplished by sealing with vacuum grease around the septa if necessary.

Set-up:

1) If not known, determine water holding capacity and moisture content of air-dried, sieved soil (Haney and Haney, 2010).
2) Weigh 10g air-dried soil into 100ml containers and place containers in canning jars. Label canning jar lids (containing septa). (*This can be done ahead of time. Rewetting the soils and baseline IRGA measurement should be done one jar at a time.*)
3) Adjust moisture to 50% water holding capacity (50% WFPS) by adding determined volume of deionized water evenly over the surface of the soil with a 5ml pipettor, or if a pipettor isn't available, use a syringe.
4) Seal jar with labeled lid (containing septa) and ring.
5) Determine a baseline CO_2 reading with the IRGA:
    a. Insert emptied syringe through the septa of incubation jar, drawing air in and out repeatedly (~ five times) to mix.
    b. Expel air in syringe until there is 0.5ml of air exactly in syringe.
    c. Remove syringe of incubation jar and inject in IRGA septa port.
    d. Record injection time.
6) Record starting weight for jar + beaker + wetted soil.
7) Place jars into 25°C incubator in the dark.

Measurements (day 0 and day 1):

1) Turn on gas tank containing IRGA carrier gas (N_2 or helium).
2) Calibrate to zero after 10 minutes of running the carrier gas.
3) Create IRGA standards in septa jars labeled A, B, and C: insert syringe needle into jar's septa to act as a vent, and then inject CO_2/N_2 mix from gas tank into jar for one minute. Remove standard jar from gas tank, then remove vent needle after jar has depressurized (listen for the air escaping from the vent needle, then remove as air stops venting).
4) Start logging measurements.
5) Take CO_2 measurements by removing 0.5ml air from the jar with a 1ml syringe and injecting it into the IRGA. Standards should be run first and last (two to three times each).
6) Record the injection time for each sample. Wait until the ppm returns to zero before entering the next sample.
7) After taking CO_2 readings, stop logging. Return sample jars to the incubator.
8) Note: on the last day of measurements, record the weights for the jars (+ beaker + soil). Any substantial weight difference from day 0 to the last day may indicate leakage.

Metric 3. Partial carbon budget

Description of the metric

As mentioned, soil carbon is a balance between the inputs of organic materials and the decomposition of SOM. A complete budget for carbon is difficult to measure or estimate, but an estimate of a partial budget would tell if similar or lower levels of carbon are going into the soil relative to that of a natural system. This can indicate whether the soil carbon levels should be increasing or decreasing compared to the natural system and among treatments. Soil disturbance such as tillage would also reduce the soil carbon relative to no- or reduced-till practices. Organic (carbon) inputs to the soil include those internal to the system (i.e., root litter and aboveground litter, including that of cover crops and the return of crop residues), and inputs external to the system (i.e., manure, composts, other forms of biomass transfer).

The scale of analysis is generally at the field level, but can be applied to the farm and landscape scales with appropriate sampling strategies and mapping.

Method of data collection and data needed

The box below provides detailed instructions for estimating a partial carbon budget.

Box 8. Method for estimating a partial carbon budget

1. Measure or ask the farmer the area of the plot which will be assessed. This is an essential step: to calculate the amount of carbon added on a rate basis, you need to know area to which it is applied.
   Size of plot_______
2. Above ground biomass inputs:
   a. Indicate or ask the farmer the primary and secondary crops (if any) that were harvested from the field in season one (and two if there is a second season). Indicate if a cover crop/green manure was grown for biomass, not grain yield. Record responses in the table below.
   b. Measure or ask the farmer the yield obtained from each crop grown in the plot and record responses in the table below.
   c. Measure or ask the farmer the biomass of the cover crop or vegetative fallow.
   d. For manure or compost calculate the amount per area (kg/ha or t/ha) of the material that was applied. Note: for each of these materials it will be necessary to estimate the percent water content and report all on a dry-weight basis. If crop residues are not measured or estimated by the farmer, then estimate the amount of residue based on the harvest index (grain/grain + biomass) for each crop.
   e. Estimate of above-ground inputs to the plot by summing the values from steps 2a, 2b, 2c, and 2d above: ___________
3. Make corrections for root inputs, organic input management, and tillage practices.
   a. Estimating below-ground inputs is extremely difficult. Obtain some type of qualitative comparison by going through the various above-ground inputs and indicating which ones would also contribute to below-ground inputs (e.g., crops, cover crops, vegetative fallows).
      Crops: roots (yes or no)
      Cover crops (yes or no)
      Other vegetative fallow (yes or no)
   b. For each yes response above, give a score of 1, otherwise a score of 0.
   c. Sum the scores from above. Total + _______.
4. Indicate or ask the farmer if the plot was managed with tillage (-1), reduced tillage (0), or no or minimum tillage (1).
5. Indicate or ask for each of the inputs if it was incorporated into the soil (yes=1, no=0) or left on the soil surface as mulch (yes=1, no=0)
   a. Crop residues (of each crop type): ________________
   b. Compost: __________________
   c. Animal manure (and what type): _______________
   d. Others: ____________________

6. Sum the numbers for root inputs, tillage practices, and inputs obtained above in steps 3, 4, and 5: ___________

Summary of and comparison of organic inputs.

1. Obtain an estimate of net primary productivity (NPP) from maps or estimates of potential yields + biomass (also from maps).
2. Compare the amount of organic inputs to the plot with that of the NPP or potential yield plus biomass.
3. Correct the comparison based on the root inputs and soil management practices calculated in 3 above by giving the number in parentheses next to the total inputs (example: if there was 5 ton/ha inputs in one treatment with a total of 4 for the modifiers, give it a 5 ton/ha (4); another plot might also have input of 5 t/ha but no modifiers to give 5 t/ha (0)
4. Compare the treatments and the NPP and give a qualitative ranking to the amount of soil carbon that would result from the treatments – NPP is usually the highest and given a 1.

Table 5. Estimating aboveground inputs for a partial carbon budget

	Crop yield	**Crop residues left after yield removed**	**Roots**
Units			
Primary crop during main season			
Secondary intercrop			
Primary crop second season			
Secondary intercrop			
Cover crop			
Other inputs	**Source (animal type or plant)**	**Amount weight or volume (include unit)**	**Wet or dry, estimate how much moisture**
Manure			
Compost			
Other amendments (describe)			

Metric 4: Earthworms

Description of the metric

The number of earthworms in a given quantity of soil is a rough indication of soil biological activity that tends to be positive for agriculture

Soil chemical quality

Description of the indicator

Soil chemical quality indicators are related to the functioning of the soil.

Metric 1. Soil pH

Description of the metric

Soil pH examines the degree of soil alkalinity or acidity. Soil pH measure the H+ concentration in the soil solution. Soil acidity, or the reaction of the soil, indicates if there are free Aluminum ions (Al+++) or Hydrogen (H+) in the soil solution. Both types of ions can affect plant growth by inhibiting root growth or diluting the other cations (i.e., nutrients such as Ca++, Mg++, K+, Na+) in the soil solution and making them less available for uptake by the plant. In general, soils with a pH in water less than 5.5 (or 6.0) are considered acidic, those with pH great than 7.2 are considered alkaline, which may also affect plant growth by making other nutrients less available. It should be noted that some plants, even crops, are more tolerant of soil acidity than others.

Measurement method: Soil test

Method of data collection and data needed

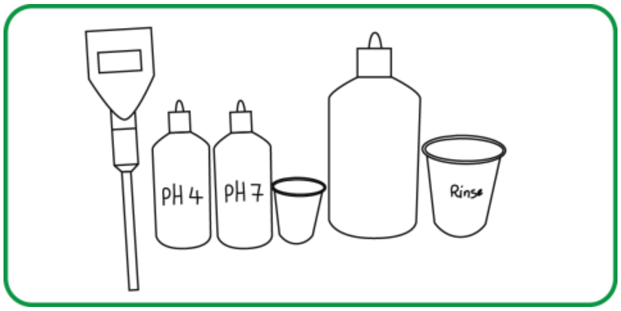
The following method has been taken directly from the SoilDoc Manual *(Excerpted from Weil and Gatere, 2016, pp 33 - 41)*

Figure 5. Calibration of the pH meter:

The calibration procedure described in the figure below should be followed. In general, calibration requires the use of two buffer standards (solutions) of known pH, bracketing the expected range of pH for the soils to be measured. For most acid soils, buffers of pH 7.0 and pH 4.0 will be adequate to calibrate the pH meter.

Materials/Instruments:

1. Hanna Instruments pH meter
2. pH meter checker (Hanna Instruments: HI 1270 pH electrode with screw-type connector: http://www.hannainstruments.co.uk/ph-electrode-for-use-with-checker1.html)
3. (2) 1.5V alkaline batteries
4. pH 4 buffer solution in 120ml dropper bottle
5. pH 7 buffer solution in 120ml dropper bottle
6. 30ml beaker
7. Centrifuge tube
8. 500ml squirt bottle filled with bottled water
9. 250ml rinse beaker



Directions for calibration:

Step 1. Remove protective black cap from pH meter electrode. Rinse pH electrode by squirting it with bottled water over the rinse beaker. Shake excess water from pH electrode. Add 10 drops of pH 7 standard solution to the cap of dropper bottle (120ml) and immerse the electrode tip into the pH 7 buffer solution in cap, making sure that the solution covers the slit up the side of the tip. Turn on the pH meter. *Do not calibrate the probe by directly measuring standard solution in dropper bottle.*

Step 2. The meter should read between 6.70 and 7.30 within a few seconds. Wait until the reading stabilizes. Adjust meter using a small screwdriver to turn the screw at top of the meter. Use very small, careful movements. Turn clockwise/counterclockwise until the LCD display reads 7.00 pH.

Step 3. Set aside cap of pH 7 solution on the vials holder (see figure below). Do NOT discard the solution.

Step 4. Shake off any drops of pH 7 buffer solution clinging to it and then rinse it thoroughly with bottled water

by squirting electrode tip over rinse beaker. Squirt bottled water up into the end of probe to flush any remaining pH 7 solution from around bulb. Shake off excess water over rinse beaker. *Rinse three times.*

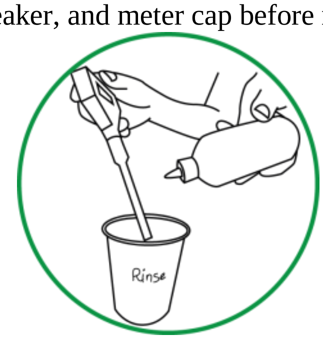
Step 5. Add 10 drops of pH 4 buffer solution into the cap of the pH 4 buffer dropper bottle and insert clean pH electrode into it. The reading on the meter should be within 3.7 to 4.3 within a few seconds. Allow the reading to stabilize. Once reading is stabilized, adjust to read 4.00 using the screwdriver on screw labeled “4/10” using small, careful movements.

Step 6. After you finish, discard the pH 4 buffer solution into a rinse beaker. Rinse out the cap and re-cap the buffer solution dropper bottle and put it in toolbox.

Step 7. Rinse the electrode as described in Step 1.

Step 8. Immerse tip of probe in pH 7 calibration solution that was aside in Step 1. The reading should be close to 7.00. Put the pH 7 solution in a glass vial, rinse the cap and re-cap the pH 7 dropper bottle.

Step 9. Rinse the pH electrode, beaker, and meter cap before moving to the next step.



Step 10. Store the pH electrode in a vial with a few drops pH 7 calibration solution until ready to sample soil solution. (Never store the meter in bottled water. Occasionally some of the storage solution will creep out and appear as dry white crystal residue on the electrode. It will dissolve when you rinse it with bottled water.)

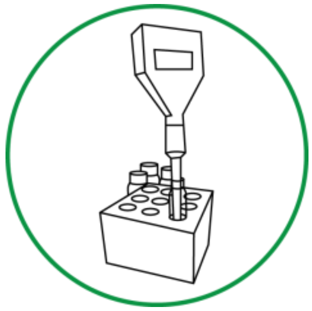


Figure 6. Measuring soil pH from a water extract

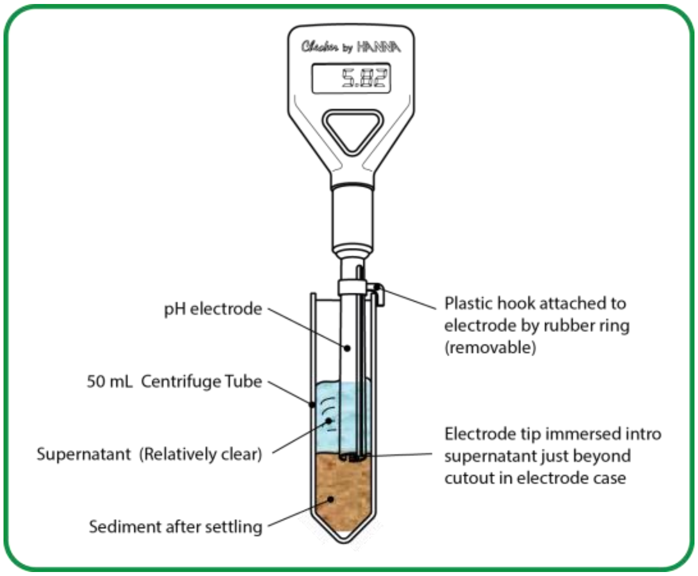
Materials/Instruments:

1. pH meter (Hanna Instruments)
2. pH meter Checker [(same as previous figure)](#)
3. Soil samples in centrifuge tubes
4. 500ml squirt bottle filled with bottled water
5. 250ml rinse beaker
6. pH electrode holder

Directions for testing pH of soil samples:

Step 1. Rinse pH electrode by squirting it with a stream of clean bottled water over rinse beaker.
Step 2. Measure and record the pH of the blank (tube 10). Rinse the pH electrode over the rinse beaker.
Step 3. For soil sample, immerse pH electrode tip into supernatant (water) above soil slurry. Stir the mixture gently.

Step 4. Use the pH electrode holder to let the meter rest on the edge of the tube so that the sensor is not resting on the soil in the tube.

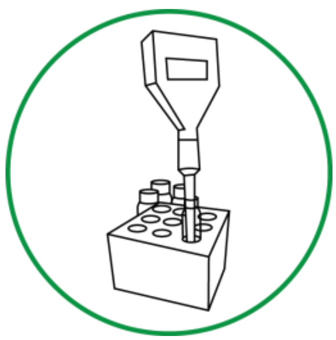


Step 5. Wait until the reading on the LCD screen stops fluctuating. (Consider the reading stable when there is no change or only up and down change for four seconds.) Record the pH. Equilibrium may be considered reached when the pH measured does not vary by more than 0.02.

Step 6. Clean the pH electrode by squirting the electrode tip with water over rinse beaker.

Step 7. Repeat Steps 3-6 for each sample.

Step 8. When finished with the pH meter, put in the vial with pH 7 solution. Don't forget to turn it off, as there is no auto-off function.



(Above method excerpted from Weil and Gatere, 2016, pp 33 - 41)

Metric 3: Electrical conductivity

Description of the metric

Electrical conductivity is an indicator for measuring soil salinity. Salinity can be a concern associated with specific soil types and with salt accrual through improper irrigation techniques.

Measurement method: Soil test

Method of data collection and data needed

The electro conductivity of the saturated paste extract is measured to determine the level of salinity.

Reagent:

Potassium Chloride

Standards:

1. Dissolve 0.7456g KCl in 1000ml water: 1.412mS/cm at 25^0C.
2. Dissolve 7.456g KCl in 1000ml water: 12.900mS/cm at 25^0C

Procedure:

1. Weigh about 300g ± 25g soil into a plastic container.
2. Add water to the soil with string until it is nearly saturated.
3. Allow the mixture to stand covered for several hours to permit the soil to imbibe the water, and then add water to achieve a uniformly saturated soil-water paste. As this point, the soil-water paste glistens as it reflects light, flows slightly when the container is tipped, slides freely and cleanly off a spatula, and consolidates easily by tapping or jarring the container after a trench is formed in the paste with the side of a spatula.
4. After mixing, allow the sample to stand (preferably overnight, but at least four hours), and then recheck the criteria for saturation. Free water should not collect on/above the soil surface, nor should the paste stiffen markedly or lose its glisten. If the paste is too wet, add additional dry soil to the paste mixture.
5. Transfer to a Buchner filter funnel fitted with Whatman No. 42 filter paper. Apply vacuum, and collect the filtrate. If the initial filtrate is turbid, refilter.
6. Measure the conductivity of the filtrate against that of the standards.

Metric 4: Soil nutrients

Description of the metric

Soil fertility and the productive capacity of the soil is dependent to a large part on the concentration of nutrients required for plant growth. There are 17 essential plant nutrients, but the macronutrients of nitrogen, phosphorus, potassium, calcium, magnesium, and sulfur are those most studied. Nitrogen (N) is the most common growth-limiting nutrient, due to constituent plant demand for nitrogen to build chlorophyll and other proteins. Phosphorus (P) is also required in substantial amounts relative to soil availability, where it is used primarily for energy-transfer molecules such as ATP. P is a major limiting nutrient, or co-limiting with N, in ancient, highly weathered soils (e.g., much of West Africa). It is important to consider soil C in relationship to N and P, as both are regulated by complex biology, and often chemistry as well. For more detail, see the articles “Nitrogen- It’s What’s for Dinner” and “Phosphorus- The Key to Sustainable Nutrient Management” from http://globalchangescience.org/eastafricanode/index.php/applied-agroecology/.

There are several approaches to assess soil nutrients:

1) Soil tests use different chemical extracts (dilute acids or bases) to measure of pool of nutrients, which are then related to “plant-available” nutrients through research. Critical values are established to indicate possible deficiencies of each nutrient that limit plant growth. There are many different soil tests that can be used.

2) Fertilizer response trials and nutrient omission trials evaluate plant growth under zero addition, and different combinations of nutrient additions.

3) Nutrient concentrations in plant tissues grown on different soil fertility levels can be compared with a given desirable range. In addition, nutrient budgets have been used to indicate if a soil is receiving more nutrients than are being lost (nutrient enrichment) or if more nutrients are being extracted through plant harvest and lost through other means (nutrient mining). Included in the additions to the soil is an assessment of the amount of nitrogen added through biological nitrogen fixation (BNF). A few of these metrics are detailed below.

It is important to keep in mind that available nutrients are a function of the plant species and rooting volume, as well as the extractable nutrient status of the soil. A shallow soil due to root restriction zones (such as a plow pan) or erosion will have a limited nutrient supply, even if soil nutrient status is high. Monitoring soil texture by depth and topography characterization can be done using simple tools. A new website provides linkages by matching geo-reference information and site characterization to on-line databases of soil profile information, providing a range of information for field practitioners (https://www.landpotential.org/). This approach is being piloted in South and East Africa, and shows considerable promise as a means to put in context the information provided by soil chemical and physical quality indicators in the next section (Herrick et al., 2016).

Metric 4. Soil nutrient levels

Method of data collection and data needed

Several methods exist for measuring soil nutrient levels that are assumed to be correlated with plant availability. These methods range from simple test strips with fertilizers applied to indicate if crops respond to addition of a nutrient, plant tissue concentration to assess if it is above or below critical values; wet chemistry, and near/mid infrared spectroscopy. The value of this last, infrared spectroscopy, as a rapid and highly cost-effective method is leading to broad use in soil survey and agricultural research, where calibration is through a subset that is analyzed using classical wet chemistry (Shepherd et al.,). The methods chosen should depend on the availability of labs or equipment to analyze the soils, human capacity to do so, cost, and the degree of accuracy needed for a particular objective.

Detailed methods are not described here. The researcher is encouraged to investigate the local labs and methods and to contract labs or people to do these analyses.

Unit of analysis

The unit of analysis for a given nutrient may vary but they are all equivalent to a measure of parts per million or parts per million of weight. For example, units used such as g/kg soil.

Limitations regarding estimation and interpretation

There are few correlations with the results from the different soil measurement methods and plant growth. Several of the methods can be costly.

Metric 5. Partial nutrient balance

Description of the metric

A partial nutrient balance estimation is recommended as fairly simple and cost-effective means for comparing if different soil management practices are depleting (negative nutrient balance) or enriching (positive nutrient balance) the soil. By comparison, a full nutrient balance calculation would consider all nutrient additions (e.g., precipitation, mineral and organic fertilizers, biological nitrogen fixation) and losses (e.g., leaching, erosion, runoff, crop harvest, gaseous emissions). Most of these variables are quite difficult and expensive to measure, so a partial nutrient balance is used to estimate the amount of nutrients added in mineral and organic fertilizers and biological nitrogen fixation (BNF) and the amount of nutrients lost through crop (or tree or animal) harvest from a given area. If the crop residues are removed from the field, then that is loss; if the residue is kept on the field, then it is not considered a loss or an addition. Only the nutrients in the grain or fruit from the harvest are considered losses for the partial nutrient balance estimation.

Unit of analysis

Partial nutrient balance estimations can be done at the field and farm level. Landscape level assessments of nutrient balance may also be done, though data are not often available and methods more difficult.

Method of data collection and data needed

Here is the method for estimation of the partial nutrient balance at the field level. Information obtained from the crop grain and residue productivity metrics and from the carbon budget metrics can be used to fill in parts of the table below. Missing data on the types and amounts of fertilizers applied can be obtained from formal or informal surveys or experimental design details.

Table 6. Worksheet for partial nutrient balance estimation

INPUTS	Amount added (kg or t/ha)	Area of field where applied and harvested	%N in materials	%P in materials	Others %nutrients	Amount of nutrient added or lost (kg/ha)
DAP						
UREA						
NPK						
Calcium Ammonium Nitrate						
Triple Super Phosphate						
Include list of more fertilizers						

Compost						
Manure						
Other organics						
BNF						

Once this table has been completed then the amount of the nutrient applied or lost can be calculated as follows:

$$Nutrients\ added\ or\ lost\ = \frac{\left((amount\ of\ material) * \left(\frac{nutrient\ concentration}{100}\right)\right)}{(area\ of\ application\ and\ harvest)}$$

The concentration of nutrients in the most commonly used fertilizers are provided in the following table.

Table 7. Percentage of nutrients in various fertilizers

Nitrogen Fertilizer	N	P_2O_5	K_2O	S	MgO
Ammonium Sulphate	21	0	0	23	0
Calcium ammonium nitrate	20.4 – 27	0	0	0	0
Urea	45-46	0	0	0	0
Single SuperPhosphate	0	16-20	0	0	0
Triple Superphospate	0	46	0	0	0
Diammonium Phosphate	18	46	0	0	0
Monoammonium Phosphate	11	52	0	0	0
NPK	5 – 25	5 - 25	5 - 25	0	0

Source: IFDC, 2012

The computation of nutrient content for fertilizers should be considered carefully. Elements like P, K, Mg that occur as oxides, it is required to use a conversion factors to convert them to the nutrients. The table below provides the conversion.

$$P_2 0_5 * 0.4364 = P$$

$$K_2O * 0.8302 = K$$

$$M_gO * 0.6030 = Mg$$

For example, if a farmer applied 100kg of diammonium phosphate (DAP), the amount of nitrogen and phosphate can be calculated as follows:

$N = (100 * 0.18) = 18$ where the percentage of N in the fertilizer (18%) times the quantity applied of 100kg gives 18kg of nitrogen applied.

For phosphate the calculation is as follows.

$P = (100 * 0.46) * 0.4364 = 20$ where we multiply the amount of P_2O_5 in DAP (46%) by the quantity applied of 100kg, then convert it to quantity of elemental phosphate with the conversion factor (0.4364), which gives 20kg phosphate applied.

The nutrient concentrations of organic materials, including crop grains, residues, manures, and composts, can be found in a variety of sources, as summarized in the following tables.

Table 8. Harvest index for given crops

Crop Residue Group	**Harvest Index**
Cereals	0.4
Sugar crops	0.56
Roots	0.4
Vegetables	0.38
Fruits	0.38
Legumes	0.49
Oil crops	0.52
Other crops	0.28

Source: Smil, 1999

Table 9. Nutrient content of harvested product and crop residues

Crop	Harvested product			Crop residue		
	Harvested product N	**Harvested product P**	**Harvested product K**	**Crop residue N**	**Crop residue P**	**Crop residue K**
	(kg/t)			(kg/t)		
Banana	1.2	0.3	4.5	1.6	0.3	11.9
Barley	15.5	2.8	6	7	1	21
Cassava	4.2	0.5	4.3	4.6	0.9	1.4
Cereals other	16.7	4.4	4.8	10.9	2.3	38.6
Citrus	1.8	0.2	2.3	0.6	0.2	4.4
Cocoa	40	8.5	19.3	19.9	4.7	33.3
Coconut	61	7.2	9.8	27	5.7	25.3
Coffee	35	2.6	16.8	4.3	3.8	9.3
Cotton	18.7	9.7	9	13.9	6	29.8
Fibres	5	0.4	6	2.1	0.7	9
Fruits other	2	0.2	2	1.8	0.2	4.9
Groundnut	37.2	6	8.2	15.9	2.4	14.9
Maize	16.8	4.1	4.8	9.7	1.9	21.4
Millet	19.2	6	5.4	20.4	4	59.8
Oil crops other	2.6	0.5	4.4	0.3	0.6	5.4
Oil-palm	2.9	0.7	4.1	3.7	0.6	3.3
Plantain	0.7	0.1	3.4	1.2	0.3	6.4
Potato	4.4	1.3	6.9	2.3	0.7	4.5
Pulses	20	3.4	11.1	10.4	1	13.1
Rice	11.6	3.4	3.4	11.3	2.3	35.8
Roots other	4.6	0.3	2.9	1.9	0.5	3.1
Rubber	6.9	1.2	4.6	1	0.2	4
Sesame	30	6.1	6.8	15	5.4	21.1
Sorghum	14.5	5.5	3.8	10.8	4.6	29.2
Soybean	62.1	10.9	20	17.6	3	14.4
Sugar cane	0.6	0.2	1.2	0.3	0.3	0.3
Sunflower	24	3.5	5.5	23	3.2	41.3
Sweet potato	4.8	0.8	7.3	2.1	1.2	3.3
Tea	35	3.8	13.4	0.1	0	0
Tobacco	56	8.2	72.7	0.1	0	0.2
Vegetables	9	0.9	2.6	3.2	1.4	7.8
Wheat	22.3	4.3	5.8	4.3	1.8	26.7

Source: FAO, 2004

The partial nutrient balance (kg or t/ha) is then estimated by:

$$Partial\ nutrient\ balance\ =\ Sum\ of\ nutrients\ added - sum\ of\ nutrients\ removed$$

If a soil is depleted (negative nutrient balance) then there may be consequence to crop growth, with some of the nutrients becoming limiting to plant growth. If the soil is fertile, then the nutrient depletion may not immediately affect plant growth, but if it continues will eventually lead to a degraded soil with reduced productive capacity. A negative nutrient balance is an indicator of unsustainable practices. If the soil is enriched (positive nutrient balance), then the level of enrichment may indicate possible levels of leaching, runoff, and/or erosion of the surplus nutrients, all of which have negative environmental impacts and indicate unsustainable practices. A neutral nutrient balance (i.e., neither positive nor negative) is the ideal and most sustainable situation.

Limitations regarding estimation and interpretation

This method is relatively easy for making comparisons among different soil and crop management practices. It provides an indication of whether the soil nutrients are being depleted or enriched. If a soil is fertile, then there may not immediately be observable consequences of depletion to soil nutrient levels and plant productivity, so this metric helps to prevent the situation of unrecognized soil nutrient depletion.

Metric 6: Biological Nitrogen Fixation

(Snapp additions below, primarily from the TSBF Handbook of Methods, appendix G by Mark Peoples, pages 164-171.)

Description of the metric

Biological nitrogen fixation is key to sustainability of farming systems, because ‘available’ nitrogen is needed in large quantities. Nitrogen is the nutrient that is most often the limiting factor for growth and production of food on smallholder farms. Legume plants and their symbiotic microorganism partners – and in particular legume food crops – can play a large role in providing much of the N through BNF. BNF provides about half of nitrogen inputs worldwide in agriculture, the Haber Bosch industrial process (manufactured fertilizer from fossil fuel feedstocks) provides the other half. There are many methods to measure symbiotic nitrogen fixation, and the most appropriate one will depend on research objectives and budget available, but this manual will focus on describing the two most important methods in wide use. This is the nitrogen difference method, and the natural ^{15}N abundance method **(**Bremer, and van Kessel, 1990)**.** If it is possible to source an iso-line of a legume species, one that doesn't have the ability to fix nitrogen, then this can be grown as the reference plant with very similar growth habit, which provides a superior reference for both methods (Kohl et al., 1980).

***Measurement** nitrogen difference method:*

Growing a non-fixer reference crop on the same site adjacent to a legume crop is one way to estimate the amount of nitrogen fixation associated with the legume crop. Both plants are grown on the same soil and exposed to identical conditions, usually be growing on adjacent plots. The additional nitrogen yield (the nitrogen content associated with the total biomass) of the legume crop compared to the reference non-fixer is used as an estimate of symbiotically fixed nitrogen. This assumes that the pattern of soil inorganic nitrogen uptake is similar for the reference and the legume plant, and any additional nitrogen accrued can thus be attributed to N fixation.

Method of data collection and data needed

One or two reference plants that are not nitrogen fixing species are planted adjacent to the legume specie(s) of interest, and a destructive harvest of above ground biomass is conducted at about maximum biomass accrual. It is important to conduct a measurement before reproduction is advanced in plant species that senesce leaves mid-season, through natural processes (e.g., shrubs such as pigeonpea *Cajanus cajan*), or due to high susceptibility to leaf pathogens (e.g., common bean). Vegetation tissue samples are collected in a representative manner from the biomass collected, and these are analyzed for nitrogen concentration. The different plant tissues can be separated if nitrogen allocation patterns are of interest. The two widely used method for tissue N determination is dry combustion of C and N (using equipment such as a Costech or Carlo Erba), and kjehdahl hot acid digestion followed by colorimetric N determination. To ascertain biological N fixation, the total amount of nitrogen in the biomass of the reference species is subtracted from the legume species biomass.

Unit of analysis

N concentration (kg N kg^{-1} biomass) multiplied by biomass mass kg ha^{-1}, final units of kg N ha^{-1}

Limitations regarding estimation and interpretation

There can be large differences in root system architecture, growth habit and plant growth rates, and it is important to use a reference plant that has a similar plant growth pattern to that of the legume crop of interest, to try and meet the assumption that both sourced from similar soil inorganic nitrogen pools and that both accumulated N in a similar manner over the season. This method is more effective in low-N fertility sites. It can in many cases provide an underestimate, as reference plants such as cereals are often used and these may have rapid growth rates and acquire N in a more rapid manner than plants with a moderate growth rate which is typical of many legumes (particularly those with a perennial growth habit).

Measurement Method 2: Natural Abundance method:

There is a small difference in the natural 15N enrichment of soil N compared to atmospheric N_2 and this can be used to quantify biological N fixation. This method is a refinement on the N difference method, as the 15N to 14N signature of soil N varies from that of the atmosphere, providing an opportunity to assess the N pools sourced by the reference plant, and the N fixer plant. Similar to the N difference method, growing a non-fixer reference crop on the same site adjacent to a legume crop is important in this methodology. In the case of established perennial plants finding an adjacent non-fixing plant to act as a reference is feasible (thus this method can be applied to situations where the N difference method is not practical to employ), as it is not necessary to collect the entire plant biomass – rather, representative plant tissue samples can be used to assess 15N to 14N signature of the non-fixer, which is expected to be closely related to the soil 15N to 14N signature. How much of legume N is derived from the atmosphere (via biological N fixation), and how much form the soil is determined using a formula that requires information about the 15N to 14N signature expected from 100% reliance on biological N fixation, and 100% on soil N.

Method of data collection and data needed

Vegetation tissue samples are collected in a representative manner from reference specie(s) and N fixer species of interest. These are analyzed for 15N to 14N signature using a mass spectrophotometer. The different plant tissues can be separated first if nitrogen allocation patterns are of interest. To ascertain biological N fixation (fNdfa), the following equation is used:

$$fNdfa = (\delta^{15}N_{ref} - \delta^{15}N_{fix})/(\delta^{15}N_{ref} - \delta^{15}N_{b})$$

where "ref" is the non-fixing plant and "fix" is the nitrogen-fixing plants grown under the same conditions, and "b" is the fixing plant grown with atmospheric N_2 as the sole external nitrogen source (Oberson et al., 2007).

Unit of analysis

Units for fNdfa: kg N kg biomass, and multiplied by biomass this is a metric for total amount of N fixation associated with this species per area: kg N ha^{-1}

Limitations regarding estimation and interpretation

Similar challenges and limitations as those observed with the N difference method except that overall growth habit differences are less important. Yet, the root system architecture does influence the ability of a reference plant to source from a similar soil inorganic nitrogen pools as the N fixer species, which is a key assumption of the natural abundance method. A mass spectrophotometer capable of precisely measuring differences of 0.00004 atom % 15N is necessary and sample preparation requires great care in order to avoid isotopic discrimination, and representative subsampling (e.g., a strategy to sample plant tissue in a representative manner, and to carry out fully homogenous sample preparation). If the soil has very low and heterogeneous soil N pools this poses a challenge; however, this is more of a problem with natural sites than most agricultural sites (Peoples, TSBF)

Soil physical quality

Description of the indicator

Soil provides the physical medium in which plants grow and roots penetrate. In addition, the physical structure of the soil allows the infiltration and storage of water and the movement of air into and out of the soil, all critical to maintaining a physical environment in which the plant grows. The physical structure of the soil also controls to a large degree the emission of greenhouse gases from the soil. Physical factors that are important to maintaining soil structure and the processes related to structure include aggregate stability and a light, non-compacted, and friable soil that has good water infiltration, gas exchange, and water-holding capacity.

Metric 1. Aggregate stability

Description of the metric

Soil particles can be held together by soil organic matter or the chemical attraction of clay particles (the stability of aggregates varies with the type of clay minerals in the soil). These aggregates contribute to the distribution of the pore space and sizes in the soil, which in turn affect the rates of water infiltration, gas exchange, and water holding capacity. Rates of water runoff and erosion are generally less in well-aggregated soil. Aggregates can fall apart as soil organic matter decomposes and as the soil is disturbed through tillage and compaction. The stability of aggregates is a function of the soil chemical, physical and biological processes.

Method of data collection and data needed

Box 9. Aggregate Stability
Aggregate stability is measured by the aggregates' resistance to slacking by water or crushing by physical pressure. It can be quantitatively assessed in the lab. If maintaining or rehabilitating soil is a major objective of a management intervention, then one might look for a laboratory that provides this analysis. There recently are more qualitative assessments of water-stable aggregates that provide a very good indicator of the relative stability of the soil under different management practices. Details for this procedure are provided below (Weil and Gatere, 2016) in a data table designed for six samples from a single soil. A receptacle sieve to assess aggregate stability is crafted from a 2mm screen attached to the bottom of a PVC tube. Two divided boxes with six compartments are used: one to hold the dry samples and one to be filled with water for the stability assessment.Follow the steps below for the different soils that will be compared.
Step 1. Fill in sample ID in the table below and select six 7-9mm diameter aggregates for the soil sample. Gently place each aggregate in a sieve and the sieve in a compartment of the dry, divided box. Make sure the samples are air-dry.
Step 2. Fill the empty (no sieves) box with water (preferably deionized, but rain or bottled drinking water will do).
Step 3. Fill each compartment to 2cm depth. The water should be approximately the same temperature as the soil.
Step 4. Test the samples in the following manner:
Step 5. Slowly lower the first sieve with its sample into the respective water-filled compartment (e.g. upper left corner of sample box to upper left corner of water box). Start timer.
Step 6. From the time the sieve screen touches the water surface to the time it rests on the bottom of the box, 1 second should elapse. Watch samples closely for 30 seconds.
Step 7. Immerse another sample every 15 seconds. Beginners may want to immerse a sample every 30 seconds. As you observe the samples, use the table to assign samples to stability classes 1 or 2.
Step 8. Observe the samples again at 5 minutes (300 sec) after they were placed in the water and record a stability class (1, 2 or 3) in the table.
Step 9. After the 5-minute observation, raise the sieve completely out of the water and then lower it to the bottom (without touching the bottom of the tray) a total of five times, taking 1 second to move the sieve in each direction (2 seconds total for each round of dipping). Do this even if you have already rated the sample a 1, 2, or 3. You should change a 1, 2, or 3 rating if >10% of soil remains on sieve after sieving. If <10% of soil remains on the sieve after five dipping cycles, repeat the rating of 1, 2, or 3 in second data row.

(Excerpted from Weil and Gatere, 2016, pp 29 – 31)

Table 2. Data table for soil stability test.

Name:	Date:			Stability rating = avg of 2nd line of scores = ________				
Sub-sample No:	1	2	3	4	5	6		
Sample/Field I.D.:								
Stopwatch time aggregate placed in water.	0:00	0:15	0:30	0:45	1:00	1:15		
Aggregate Stability Class (1-3)								
Stopwatch time start to dip in and out of water 5 times.	5:00	5:15	5:30	5:45	6:00	6:15		
Aggregate Stability Class (1-6)*								

Table 3. Stability class descriptions

Stability Class	**Criteria for assignment to stability class (not counting gravel)**
1	50% of structural integrity lost (melts) within 5 seconds of immersion in water, OR soil too unstable to sample (falls through sieve).
2	50% of structural integrity lost (melts) 5-30 seconds after immersion.
3	50% of structural integrity lost (melts) 30-300 seconds after immersion, OR < 10% of soil remains on the sieve after five dipping cycles.
4	10–25% of soil remains on the sieve after five dipping cycles.
5	25–75% of soil remains on the sieve after five dipping cycles.
6	75–100 % of soil remains on the sieve after five dipping cycles.
Note: Hydrophobic samples (float in water after pushed under) are rated 6. Gravels >2mm that were part of the original sample need to be subtracted from both the original volume and volume remaining on the sieve after 5 dips. If gravel is suspect, attempt to rub remaining sample through sieve with finger to determine if gravel is present.	

Figure 4. Illustration of soil stability classes. Note that actual soil clod or aggregates should be about 1/4 inch in diameter (Herrick, et al., 2005).

Metric 2. Bulk density

Description of the metric

Bulk density is a measure of soil compaction, which may affect water holding capacity, infiltration, and nutrient availability (Doran and Jones, 1996).

Measurement method: Field and soil tests

Method of data collection and data needed

Measuring bulk density requires soil water content that is at normal field capacity. Field capacity is the soil moisture or water content held in the soil after drainage of excess water. Soil to be used for analysis should not be very dry and a small deviation from field capacity will not bias the results (Anderson and Ingram, 1993; Doran and Jones, 1996).

Procedure for measuring bulk density for non-stony soils:

1. Remove 1-2 cm of surface soil from the spot where samples will be taken and level the spot.
2. Drive a 5cm diameter thin sheet-metal tube of known weight (B) and volume (V) 5cm into the soil surface.
3. Excavate the soil from around the tube and cut the soil beneath the tube bottom.
4. Trim excess soil from the tube ends.
5. Dry at 105°C for 2 day, and weigh (C)

Unit of analysis

The unit of analysis of bulk density is g/cm^3. To calculate bulk density from the above measurements (Anderson and Ingram, 1993):

$$Bulk\ density\ (g/cm^3) = (B - C)\ /\ V$$

(Excerpted from Anderson and Ingram, 1993, pp. 95)

Metric 3. Water-holding capacity

Description of the metric

Water-holding capacity is the amount of water that a soil horizon can store for plant use. It is estimated as the difference between lower limit of plant available water and the field capacity.

Measurement method: Field and lab tests

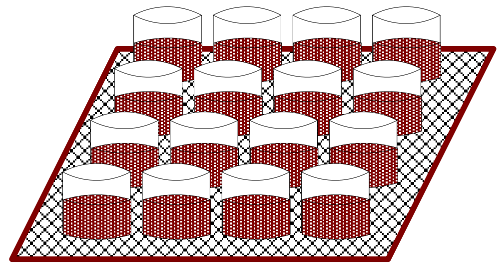
Method of data collection and data needed

Soil Water-Holding Capacity Determination

Reference:

Canadian Soil Methods book

Equipment and Materials:



Balance (capacity and precision)

Specimen cups – several small holes drilled in bottom perimeter

Milk filters, cut to fit inside bottom of specimen cups

Large plastic containers

Screen drain platforms

Small sample tins

Reagents:

Distilled water (DH_2O)

Method:

1. Allow soil samples to remain open to air until dry. Usually 72 to 120 hours depending on moisture content and container size.
2. Air-dried soil should be coarsely sieved through 5 mm sieve. Discard materials remaining in sieve.
3. Moisten milk filter with DH_2O and place it over the bottom of the perforated specimen cup.
4. Weigh the wet, empty specimen cup with filter and record the weight and cup number
5. Weigh approximately 150g of sieved soil into the moistened cup and record the total weight.
   a. Simultaneously, weigh approximately 10-20 g of sieved soil into a numbered sample tin. Record the empty tin weight, the sample weight and tin number.
   b. Place these samples in a forced-air drying oven at 100 ºF for at least 12 hours.
   c. After a minimum of 12 hours, remove dried samples from the oven and weigh. Record dry weights of the tin + sample.
6. When all specimen cups are filled and weighed, arrange them on screen drains inside the large plastic containers.
7. Add DH_2O to the bottom of the large container until the bottoms of all specimen cups are submerged by at least 1 cm. Leave the cups in the water until the soil on each top surface is moist and glistening.
8. Remove the screen drain with sample cups to an air-tight container. Close tightly. Allow cups to drain for 48 hours.
9. After 48 hours, open the air-tight container and remove excess water from the bottom of each sample cup with a sponge. Weigh each wet sample cup and record the weight.
10. Calculate moisture content of the air-dried sample and the water-holding capacity for each sample.

Calculations:

Moisture Content of Air-dry Sample (%) =

$$\frac{[(air\ dry\ soil + tin) - empty\ tin]}{[(oven\ dry\ soil + tin) - empty\ tin]} * 100$$

Water-Holding Capacity (g H2O/g soil) =

$$\frac{[(moist\ soil + cup) - empty\ cup]}{[(air\ dry\ soil + cup) - empty\ cup]} * [1 - moisture\ content\ of\ \ air\ dry\ sample]$$

Sources:

Milk filters # D547 - Ken Ag Milk Filter Tube for pipeline systems (4-7/8" x 17") 50 per box

American Livestock Inc. Madison, WI [www.americanlivestock.com]

Specimen cups #15310800 - 4 oz. plastic non-sterile polypropylene with screw cap

MSU University Stores

Metric 4: Water infiltration rate

Description of the metric

Infiltration rate is a measure of the speed at which water moves into soil, i.e., how rapidly it enters the soil. A soil with slow infiltration leads to water ponding and soil saturation, which deprives crops of oxygen. On a sloped site, slow infiltration leads to erosion from surface runoff of water that can't soak in. Good infiltration is important for capture and retention of water and to recharge soil moisture.

Measurement method: Ring infiltrometer

Method of data collection and data needed

A ring infiltrometer is the simplest method to measure water infiltration in the field. In the field, improved measurements can be obtained using a sprinkler infiltrometer (Thierfelder and Wall, 2009) or the stick method (https://jornada.nmsu.edu/monit-assess/manuals/monitoring).

As well, there are lab methods using intact soil cores; however, these methods are complex and time consuming. It is important to be aware the soil surface infiltration tends to be highly variable, as the surface roughness and pore distribution is often very heterogeneous. Obtaining multiple infiltration measurements across a field site is one way to address this challenge, suggesting a simple, relatively rapid method is useful as it allows more replications to be completed.

Before beginning, practice driving the ring infiltrometer into the soil without disturbing the soil surface. Using the graduated cylinder, pour water into the infiltrometer and note that the water "ponds" within the ring. Select a penetration depth and ponding level that will be maintained for each repetition of the measurement. Recommendations: A 12.5cm diameter ring, a 3cm installation depth into soil, and water added to a depth of 3cm in the ring.

1. Select a representative site, one that has no large cracks. Trim any vegetation close to the surface and move aside any residues for a clear surface. Pre-wet the soil to a depth of 4cm by laying a pre-moistened cloth towel on the surface and then pouring on – slowly, in multiple small amounts – a cup of water (about 370ml). Wait a few minutes then apply a second cup in the same manner (an additional 370ml).

2. Drive the infiltrometer ring into the soil to a depth of 3cm (mark the outside). A beveled edge on the ring will help drive it in smoothly, and a piece of wood can be used to evenly distribute the pressure as it is installed. You also can slightly twist it to get it to go into the soil. Test that the ring is securely set in the

soil (if it moves around, then insert it another 0.5 cm).

3. Checking for leaks and pre-wetting: Add water into the ring without disturbing the surface. This can be accomplished by laying plastic wrap on top of the infiltration ring and slowly removing to allow uniform release of water. Pour about 3cm depth water into the ring (370ml water for a 12.5 cm diameter ring). Check for any signs of visual leaks, if leaks appear, then push the ring in another 0.5 cm. Distilled water or rain water are preferred.

4. Measuring infiltration rate: Add another 370 millimeters water to the infiltrometer ring (using the plastic wrap and careful removal as above). Start a stopwatch at the exact moment the wrap is removed and the water is allowed to start infiltrating. Record how long it takes in seconds for the water to infiltrate until about 50% of the soil no longer glistens (is shiny). This is the time it takes for 3cm of water to infiltrate into the soil. This is recorded as mm/min.

Figure 5. Illustration of use of plastic wrap to hold water in ring infiltrometer before allowing water to flow into the ring for timing of infiltration.

Unit of analysis and calculations

To compute infiltration rate, convert the volume of water to a water depth, then divide by the elapsed time it takes for water to completely infiltrate (50% of soil glistens). For example, if it takes 10 minutes for 370ml (3cm depth) of water to infiltrate, then to report the infiltration as cm/sec:

$$3\ cm\ /\ (10\ min\ *\ 60\ sec/min)\ =\ 0.05\ cm/sec$$

To calculate depth of water:

1. A = surface area of the infiltrometer ring. Where radius (r) = half of the ring diameter in centimeters, and π can be looked up in a table or on a calculator.

$$A = \pi r^2$$

2. Calculate the depth of water infiltrated (H) as the volume of water (V) divided by the surface area (A) of the infiltrometer, where water volume is measured in milliliters (100ml added = 100cm^3 because 1 ml = 1 cm^3.

$$H\ = V/A$$

3. Record the time elapsed in seconds, and calculate the infiltration rate (I) as water depth (H) by time elapsed (t in cm/sec):

$$I = H/t$$

Limitations regarding estimation and interpretation

If the soil is already saturated (flooded) and it is not possible to carry out water infiltration, the soil will have to be allowed to dry for several days before testing infiltration.

This method cannot be used in very sandy, stony, or clayey soils, or on frozen ground. It is important for the ring to not leak, nor to be set over large pores or cracks.

Carrying out multiple infiltration measurements over an area will improve the assessment, as infiltration rate tends to vary markedly over time and space. Using a transect method to select measurement sites and a bottle and pipette to improve test accuracy have been shown to improve the reproducibility as a more quantitative method. However, it is time consuming and the semi-quantitative bottleless method described above used over many locations often provides a more simple and comprehensive description of water infiltration properties (Herrick et al., 2005 or https://jornada.nmsu.edu/monit-assess/manuals/monitoring)

Greenhouse gas emissions

Description of the indicator

We do not recommend this indicator.

Pesticide use

Description of the indicator

Pesticide use indicators focus on aggregate active ingredients in biocide compounds – such as insecticide, herbicides, fungicides, nematicides – that may impact human health, water quality, and lead to death or extinction of species (Aktar et al., 2009; Padovani et al., 2004). Although biocides are used to increase crop productivity by reducing or killing weeds, insects, rodents, and other organisms, the active ingredients in any type of biocide can also have unintended consequences to water quality, biodiversity, and human health. For an in-depth discussion of pesticides and contamination, refer to the Pesticide Contamination Metric in the Human Condition Domain.

Metric 1: Active ingredient applied per hectare

Description of the indicator

Active ingredients and the concentration of the biocide can provide an indication of the level of exposure for humans, crops, and insects. For example, the concentration of active ingredients in biocides ranges from 2% to 80%. By comparison, household-use pesticide concentrations are in the 2% range, which purposely limits the levels of exposure (WHO, 2015).

Measurement method: Agricultural survey (recall)

Method of data collection and data needed to compute metric

Recall surveys are used to collect data on pesticide use for agricultural production. The farmers are asked if any pesticides were used, and the amount and unit of measure (e.g., liter, kg). In addition, surveys can be modified to collect data on the type of pesticide used. This is something that most multi-indicator surveys (e.g., LSMS) do not ask. If data on the type of pesticide is not collected, then it is difficult to know the concentration of active ingredients applied to the farm. In addition, data on the area (in acres or hectares) on which pesticides are applied should be collected.

Unit of analysis

The unit of analysis is the active ingredient that is applied per hectare. WHO (2002) provides a classification of active ingredients by physical state (solid or liquid) and their level of hazard.

Limitations regarding estimation and interpretation

An issue with this metric is the recall period in which the farmer has to provide accurate data on pesticide use. In most cases, surveys collect data at the end of the season or agricultural year, and the information depends on the correct recollection of information by the farmer(s).

References

Asian Development Bank (ADB). 2016. Asian Development Outlook 2016. Strengthening Water Security in Asia and the Pacific. https://www.adb.org/sites/default/files/publication/189411/awdo-2016.pdf

Bonjour, S., Adair-Rohani, H., Wolf, J., Bruce, N.G., Mehta, S., Prüss-Ustün, A., Lahiff, M., Rehfuess, E.A., Mishra, V. and Smith, K.R., 2013. Solid fuel use for household cooking: country and regional estimates for 1980-2010. *Environmental Health Perspectives (Online)*, *121*(7), p.784.

Bremer, E. and Van Kessel, C., 1990. Appraisal of the nitrogen-15 natural-abundance method for quantifying dinitrogen fixation. Soil Science Society of America Journal, 54(2), pp.404-411.

Brocard, D., Lacaux, J.P. and Eva, H., 1998. Domestic biomass combustion and associated atmospheric emissions in West Africa. *Global Biogeochemical Cycles*, *12*(1), pp.127-139.

Brouwer, I.D., Hoorweg, J.C. and Van Liere, M.J., 1997. When households run out of fuel: responses of rural households to decreasing fuelwood availability, Ntcheu District, Malawi. World development, 25(2), pp.255-266.

Cline-Cole, R.A., Main, H.A.C. and Nichol, J.E., 1990. On fuelwood consumption, population dynamics and deforestation in Africa. *World Development*, *18*(4), pp.513-527.

Culman, S.W., S.S. Snapp, L.E. Gentry and J. Green. 2013. Short and long-term dynamics of labile soil C and N pools reflect management and predict corn agronomic performance. Agronomy J. 105:493-502.

Ejigu, M., 2008, May. Toward energy and livelihoods security in Africa: Smallholder production and processing of bioenergy as a strategy. In Natural Resources Forum (Vol. 32, No. 2, pp. 152-162). Blackwell Publishing Ltd.

Food and Agriculture Organization of the UN (FAO). 2004. Scaling Soil Nutrient Balances: Enabling mesolevel http://www.fao.org/docrep/008/y5749e/y5749e09.htm

Franzluebbers, A., R. Haney, C. Honeycutt, H. Schomberg, and F. Hons. 2000. Flush of carbon dioxide following rewetting of dried soil relates to active organic pools. Soil Sci. Soc. Am. J. 64:613–623.

Haney, R.L., Brinton, W.H. and Evans, E., 2008. Estimating soil carbon, nitrogen, and phosphorus mineralization from short ‐ term carbon dioxide respiration. Communications in Soil Science and Plant Analysis, 39(17-18), pp.2706-2720.

Haney, R.L., Haney, E.B. 2010. Simple and rapid laboratory method for rewetting dry soil for incubations. Communications in Soil Science and Plant Analysis. 41(12):1493-1501.

Haney, R.L., Haney, E.B. 2010. Simple and rapid laboratory method for rewetting dry soil for incubations. Communications in Soil Science and Plant Analysis. 41(12):1493-1501.

Herrick JE. 2005 (stablility analysis cited – need reference)

Herrick JE. et al. 2016 The land potential knowledge system (LandPKS): mobile apps and collaboration for optimizing climate change investments. Ecosystem Health and Sustainability

Herrick, J.E., Van Zee, J.W., Havstad, K.M., Burkett, L.M. and Whitford, W.G., 2005. Monitoring manual for grassland, shrubland and savanna ecosystems. Volume I: Quick Start. Volume II: Design, supplementary methods and interpretation. Monitoring manual for grassland, shrubland and savanna ecosystems. Volume I: Quick Start. Volume II: Design, supplementary methods and interpretation.

Herrick, J. E., Justin W. Van Zee, Kris M. Havstad, Laura M. Burkett, and Walter G. Whitford. "Monitoring manual for grassland, shrubland and savanna ecosystems. Volume I: Quick Start. Volume II: Design, supplementary methods and interpretation." *Monitoring manual for grassland, shrubland and savanna ecosystems. Volume I: Quick Start. Volume II: Design, supplementary methods and interpretation.* (2005).

IFDC, 2012 Training Manual on Fertilizer Statistics in Africa. http://www.inter-reseaux.org/IMG/pdf/AfricaFertilizer-org_-_Training_Manual_on_Fertilizer_Statistics_-_english_June_2012_.pdf

Kohl, D.H., Shearer, G. and Harper, J.E., 1980. Estimates of N2 fixation based on differences in the natural abundance of 15N in nodulating and nonnodulating isolines of soybeans. Plant Physiology, 66(1), pp.61-65.

Markwell, J., Osterman, J.C. & Mitchell, J.L. 1995. Calibration of the Minolta SPAD-502 leaf chlorophyll meter. Photosynth Res 46: 467. doi:10.1007/BF00032301

McBride, A.C., Dale, V.H., Baskaran, L.M., Downing, M.E., Eaton, L.M., Efroymson, R.A., Garten, C.T., Kline, K.L., Jager, H.I., Mulholland, P.J. and Parish, E.S., 2011. Indicators to support environmental sustainability of bioenergy systems. Ecological Indicators, 11(5), pp.1277-1289.

Pate, J.S., Unkovich, M.J., Armstrong, E.L. and Sanford, P., 1994. Selection of reference plants for 15N natural abundance assessment of N2 fixation by crop and pasture legumes in south-west Australia. Australian Journal of Agricultural Research, 45(1), pp.133-147.

Peoples, M. TSBF Handbook of Methods, appendix G, pages 164-171.

Orr, A., Kambombo, B., Roth, C., Harris, D. and Doyle, V., 2015. Adoption of integrated food-energy systems: improved cookstoves and pigeonpea in southern Malawi. *Experimental Agriculture*, *51*(2), p.191.

Renard, K.G., Foster, G.R., Weesies, G.A. and Porter, J.P., 1991. RUSLE: revised universal soil loss equation. Journal of soil and Water Conservation, 46(1), pp.30-33.

Smil, V., 1999. Crop Residues: Agriculture's Largest Harvest Crop residues incorporate more than half of the world's agricultural phytomass. Bioscience, 49(4), pp.299-308.

Snapp, S.S. and V.L. Morrone 2008. Soil quality assessment. Chapter 7 *In*: S. Logsdon (Ed.) 'Soil Science: A Step-by-Step Analysis' American Society of Agronomy, Soil Science Society of America, Madison, WI.

Tuomisto, H., 2010. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography*, *33*(1), pp.2-22.

U.S. EPA (United States Environmental Protection Agency). 2005. National estuary program—challenges facing our estuaries. Key management issues: Nutrient overloading. https://www.epa.gov/nep/how-national-estuary-programs-address-environmental-issues

Vital Signs. 2014a. E-Plot Biomass Measurements. http://vitalsigns.org/sites/default/files/VS_Protocol_EPlots_August2016.pdf

Vital Signs. 2014b. E-Plot Soil Sampling and Processing. http://vitalsigns.org/sites/default/files/VS_Protocol_Soils_August2016.pdf

Weil RR, Islam KR, Stine MA, Gruver JB, Samson-Liebig SE. 2003. Estimating active carbon for soil quality assessment: A simplified method for laboratory and field use. American Journal of Alternative Agriculture 18:3–17.

Weil, R.R., and Gatere, L. 2016 SoilDoc Kit System. SoilDoc Protocols

Wischmeier, W.H. and Smith, D.D., 1978. Predicting rainfall erosion losses-a guide to conservation planning. Predicting rainfall erosion losses-a guide to conservation planning.

World Health Organization (WHO) 2002. The WHO Recommended Classification of Pesticides by Hazard and Guidelines to Classification 200-2002. http://www.who.int/ipcs/publications/en/pesticides_hazard.pdf (Accessed 8/22/17)

Appendix A

50% Water-Filled Pore Space (WFPS) for CO_2 Respiration Assay

This procedure describes a gravimetric method for determining the appropriate amount of water to rewet air-dried soils to 50% WFPS in preparation for laboratory incubation. The percentage of soil pore space filled with water, as determined by water content and total porosity, is closely related to soil microbial activity (Linn and Doran, 1984). Current literature indicates that a range of 30 to 70% WFPS is sufficient for peak microbial activity (Haney and Haney, 2010).

Materials:

- 50-ml plastic beakers with three to five small (~6mm) drainage holes drilled in the bottom
- 47mm Whatman filters
- Half-pint canning jars

Procedure:

1. Add filters to the bottom of 50ml plastic beakers with small drainage holes.
2. Record the weight of the beaker + filter before filling with soil.
3. Weigh 40g soil into graduated cylinder to determine volume.
4. Transfer soil into beakers.
5. Place beakers of soil in canning jars (beakers should sit on lip of jar, allowing excess water to drain.)
6. Add 30ml water to each beaker.
7. Leave to drain for 24 hours.
8. After 24 hours, weigh beakers to determine the wet weight of the sample.
9. Dry @ 105°C for 24 hours.
10. After 24 hours, weigh beakers to determine the oven-dried weight of the sample.

Calculations:

$$Volumetric\ water\ content\ (g/g) = \frac{[Weight\ of\ moist\ soil - Weight\ of\ dried\ soil]}{Weight\ of\ dried\ soil}$$

$$Soil\ bulk\ density\ (g/cm^3)\ =\ Oven-dried\ weight\ of\ soil\ /\ Volume\ of\ soil$$

$$Soil\ porosity\ =\ 1\ -\ Soil\ bulk\ density\ /\ Particle\ density\ (assumed\ to\ be\ 2.65\ g/cm^3)$$

$$Volumetric\ water\ content\ (g/cm^3)\ =\ Soil\ water\ content\ \times\ Bulk\ density$$

$$WFPS\ (\%)\ =\ Volumetric\ water\ content\ \times\ 100\ /\ Soil\ porosity$$

In summary, WFPS can be determined using the following equation:

$$WFPS\ = \frac{Volumetric\ water\ content\ \times\ Bulk\ density}{[1\ -\ Bulk\ density\ /\ Particle\ density]}$$

or

$$WFPS\ (\%) = \frac{\left[\frac{(Weight\ of\ moist\ soil - weight\ of\ dried\ soil)}{volume\ of\ soil}\right]}{[1 - (weight\ of\ dried\ soil\ /\ volume\ of\ soil)\ /\ 2.65]} * 100$$

Human Condition Domain (Note: The superscript letters ([a,b,c]) after each metric refer to the methods in the right-hand column)

HUMAN CONDITION DOMAIN					
Indicator	**Field/plot level**	**Farm level**	**Household level metrics**	**Community/ Landscape + metrics**	**Measurement method**
Nutrition	Protein production (g/ha) [a,b] Micronutrient production (g/ha) [a,b]	Total protein production (g/ha) [a,b] Total micronutrient production (g/ha) [a,b] Availability of diverse food crops [a]	Access to nutritious foods [a] Dietary diversity [a] Food consumption score [a] Nutritional status (underweight, stunting, wasting) [c] Uptake of essential nutrients [d]	Market/landscape supply of diverse food [a,e] Dietary diversity [a] Rate of underweight, stunting and wasting [c] Average birthweight [c]	[a] Survey [b] Look up tables [c] Anthropometric measurements [d] Blood tests [e] Participatory mapping
Food security	Food production (Calories/ha/year) [a,b]	Food production (Calories/ha/year) [a,b]	Food availability [a] Food accessibility [a] Food utilization [a] Food security composite index [a] Months of food insecurity [a] Rating of food security [c]	Total food production [a] % population food secure	[a] Survey [b] Look up tables [c] Participatory assessment
Food safety			Biological contaminants Mycotoxins (toxicity units per gram) [a] Chemical contaminants Pesticide contamination [a,b] Heavy metal contamination [a] Physical contaminants Quantity of rocks per ton of grain [c]	Incidence of food borne diseases (E.coli, Salmonella, Campylobacter)	[a] Laboratory testing [b] Health center data [c] Sorting and weighing
Human health				Incidence of zoonotic diseases [a] Incidence of vector borne diseases [a]	[a] Health center data
Capacity to experiment			# of new practices being tested [a,b]	% of farmers experimenting [a,b]	[a] Individual survey [b] Focus group

Human Condition Domain

Nutrition

Description of indicator

Nutrition[5] is both an output and input for sustainable agriculture. The choices of what foods to produce, market, and consume have direct effects on nutritional outcomes. Good nutrition plays an important role in achieving the optimal childhood development and supply adults with proper nutrition to be productive individuals (UNSCN, 2015). Production and consumption of nutritious food may alleviate the burden of undernutrition, overweight, and micronutrient malnutrition at the household and individual levels (IFPRI, 2014) that are key components of sustainable development. Nutritional and dietary quality indicators focus mainly on women and young children who are the groups most vulnerable to malnutrition. A more common measure of nutrition at the household and community scale is the use of anthropometric measurements. There is a focus on increasing the access to nutritious diets or foods through nutritional-sensitive agriculture[6]. Examining nutritional outcomes from agricultural interventions is challenging and should be done with consultation with a nutrition expert.

Metric 1: Protein production

Description of metric

Consumption of protein in the diet is essential for growth and maintenance of the human body and, in combination with other micronutrients, may reduce incidents of undernourishment among vulnerable populations in agrarian economies. Protein is also a key nutrient especially in the first 1000 days of life of an individual where consumption of enough protein may prevent wasting and stunting (WFP, 2015). This indicator is used to assess the potential availability of protein from an intervention at the field level. It may be aggregated to farm/household level if the units of analysis are standardized for computation. But caution should be taken in interpretation of this indicator. It provides information on the "potential availability" of protein to the household. This metric does not provide information on the impact that this intervention will have on the nutrition of the individual or household. Interpretation should be done with caution to avoid providing inaccurate or incomplete information.

Method of data collection and data needed to compute metric

Data to compute this measure should come from two sources. First a survey is used to collect data on the agricultural output from using a given technology (crop or animal productivity indicator section), and then food composition tables are used to determine the amount of protein contained in a kilogram of the given product. It is important to note that the food composition tables provide data for each food product. For example, the amount of protein in maize per 100 grams will be different from that in beans per 100 grams or meat (FAO, 2017). Use of food consumption tables that are country specific (if available) may provide optimal information on the food content for that country.

[5] "There is no automatic mechanism in which agricultural projects positively impact nutrition, but there are plenty of entry points if one carefully designs those projects in a nutrition sensitive manner."

[6] Maximize agriculture's contribution to nutrition

Unit of analysis

The unit for analysis for this indicator is grams of protein per hectare. Using the nutrient composition tables (FAO, 2017), the value may be calculated as (note that the value should be in grams of protein per 100g for that product);

> $PC_i * 10 * YD_i$ where PC is the protein content be 100 grams of crop or product "i" and YD is the yield, in kilograms per hectare (see crop and animal productivity indicators) for given crop of animal product or by product.

Limitation

This indicator examines the potential availability of micronutrients for consumption. It does not tell us much about the impact of the intervention on nutrition of the individual, which would require observed or measured consumption. In addition, nutrition assessment is complex and requires, among other things, measurement of consumption and bioavailability, i.e., proportion of the nutrient that is absorbed or metabolized by the body through normal functions (De pee and Bloem, 2007). Bioavailability may also be affected by crop preparation and cooking practices that may limit the nutrient available that cannot be captured by this metric. Since these aspects are not observed, the indicator is a proxy of potential nutrient contribution and must be treated with caution. In cases, where the technology compared new cultivars that are biofortified, the food composition tables may not provide an accurate estimate of nutrients in that variety. The scientist may explore some laboratory tests of the composition or may ask the breeder about the composition of the new variety. In addition, once the variety is planted in different locations, the micronutrients in the output may differ due to the biophysical factors. These potential variations should be accounted for.

Metric 2: Micronutrient production

Description of metric

Micronutrients, mainly vitamins and minerals, are nutrients that are required by the human body to carry out physiological functions (Burchi et al., 2011). There are 19 identified essential micronutrients that are critical for optimal immune system functioning, physical and mental development, and metabolic processes (Kennedy et al., 2007). A person may have adequate calories but may be lacking in essential nutrients ("hidden hunger"), and though the signs of these deficiencies are not visible, the long-term effects on health, physical and mental development, and productivity may be devastating (Burchi et al., 2011; Muthayya et al., 2013). Addressing such deficiencies through nutritional sensitive agriculture may be a key issue in providing nutrients to the targeted population. Micronutrient deficiencies may be context and population specific, but the most widespread micronutrient deficiencies addressed in past projects have been vitamin A, iron, and iodine (Allen, 2000). For example, vitamin A adequacy for infants (0-5 years) can reduce infections of malaria, measles, and diarrhea; and iron adequacy can reduce anemia, which affects more than 25% of the world's population, in particular infants and women (WFP, 2015). This metric mainly examines the "potential" availability of nutrients to the producer from the crops or animal products produced.

Method of data collection and data needed to compute metric

As indicated in the productivity domain and food security sections, data on crop and animal production plus nutrient composition tables are needed to calculate this metric. The latter provides an estimate of the micronutrients that are available in given crop or food product. The Food and Agricultural Organization (FAO, 2017) and some universities (HSPH, 2016) have developed composition tables for various foods some with additional country specific food composition tables. When not cost prohibitive, a more precise measure of particular nutrient composition may be done in the laboratory for a given technology.

Unit of analysis

The unit of analysis is grams (g) of the micronutrients produced per hectare. Data on the micronutrients is provided as grams of micronutrient unit per 100 g for some but not all nutrients (FAO, 2017). In cases where composition is not as gams of micronutrients per 100 g, an appropriate conversion to grams should be made. For example, iodine is usually reported in micrograms (mcg) and iron in milligrams (mg) per 100 g of the product. Caution must be taken in converting these values to the appropriate grams per 100 g of product. Once the values are in micronutrients per 100 g of product, multiply the micronutrients per 100 g by the yield (kilograms per hectares from productivity domain) and then by 10 (to obtain the amount of micronutrients in a kilogram).

Limitations

This limitation is similar to metric 1. This indicator examines the potential availability of micronutrients for consumption and does not provide information on whether the product will or was consumed by the household. . In addition, nutrition assessment is complex and requires, among other things, measurement of consumption and bioavailability, i.e., proportion of the nutrient that is absorbed or metabolized by the body through normal functions (De pee and Bloem, 2007). Since these aspects are not observed, the indicator is a proxy of potential nutrient contribution and must be treated with caution."

Metrics 3 and 4: Total protein production and total micronutrient production

The metrics total protein production and total micronutrient production at the farm level can be constructed using same methods at the field level. The macro or micronutrients must be aggregated across the field for the respective interventions and divided by the number of hectares used to produce the output.

Metric 5 - Dietary diversity score

Description of metric

Dietary diversity is a measure that examines the variety of foods consumed at the individual and households levels. At the individual level, the score may be used as proxy for nutritional quality of the diet, and at the household level the score is indicative of access and consumption of diverse macro and micronutrients (FAO, 2013). In addition, the household dietary diversity score can be used as a proxy measure for household food access based on the following assumptions: 1) a more diversified diet is associated with improved outcomes such as birth weight, hemoglobin concentrations, and child anthropometric status; 2) diversified diet is correlated to caloric and protein adequacy, household income, and high quality protein (animal); 3) this metric can be assessed at the household and individual level to allow for household and intrahousehold assessment; and 4) data collection to estimate this metric is easy to implement as a survey (Swindale and Bilisky, 2006).

Method of data collection and data needed to compute metric

Data for this measure is collected at the household level via survey and is implemented at the individual or household level (questions about individual or household consumption). In this case, we focus on the household consumption. The survey is best implemented in a period of the year with greatest food deficiency (for example, lean months prior to harvest). If the similar survey is to be performed the next year, it is best performed at the same time of the year in order to have relevantly comparable data. The questions should be asked to the person responsible for food preparation. A 24-hour reference period (1 day)[7] is recommended because longer recall periods may lead to inaccuracies. Data should be to determine if household members consumed food from the 12 food groups (Table 3). (Note that the table provides both groups that can be used for the household and individual surveys.)

Table 3: Food groups that are considered in at the household and individual level

Household dietary diversity score food groups (score: 0-12)	Individual dietary diversity score (children) food groups (Score:0-8)
Cereals	Grain, roots, or tubers
Roots and tubers	Vitamin A-rich plant foods
Vegetables	Other fruits or vegetables
Fruits	Meat, poultry, fish, seafood
Meat, Poultry, offal	Eggs
Eggs	Pulses/legumes/nuts
Fish and seafood	Milk and milk products
Pulse/legumes/nuts	Food cooked in oil/fat
Milk and milk products	
Oil/fats	
Sugar/honey	
Miscellaneous	

Source: Swindale and Bilisky (2006)

Questions should be asked so that household food preparer responds "yes" or "no." Please note that for given contexts, one also should ask about locally consumed foods and map them to the food groups listed. For example, millet is not specifically on the list but may be a food crop in some settings, it should be included in the cereal list.

Unit of analysis

Once the data is collected, the calculation of the score is fairly straightforward. If within the last 24 hours, the household consumed maize, then for cereal a value of 1 is assigned if not then 0. Once all the values are assigned per group, sum the values across the 12 groups. If a food item from each food group was consumed, then the max summed value is 12 and if none then 0. The higher the value, the more diversified the diet and more likely the household is to have a nutritious diet.

[7] Some studies have chosen longer recall periods like 4 weeks to deal with other issues like seasonality and survey intensity (Hammond et al., 2016)

Metric 6: Food consumption score

Description of metric

The food consumption score is used to examine prevalence of food security. It uses data from dietary diversity and household food access over a seven-day recall period. This indicator has been widely used by the World Food Program (WFP, 2008; Jones et al., 2013). The diversity of consumption is analyzed across a given number of food groups, normally eight, with assigned weights. Modification to the weights may be done depending on the objective of the study and the context (WFP, 2015). Such modification may be done in consultation with an expert in this field.

Method of data collection and data needed to compute metric

Data for calculating the food consumption score is collected at the household level using a survey. A seven-day recall is period is used to enumerate the number of times the household consumed a given food. A list of items per food group should be provided. In addition, standard weights are available (see Table X), but there are methods in the literature that are available if the researcher would like to compute custom own weights (WFP, 2008). The calculation steps for the food consumption score are as follows (WFP, 2008):

1. Using the food frequency data, group all food items into their specific groups (see Table X)
2. Sum the consumption frequencies of the food items in the same group and recode the values of each group greater than 7 as 7.
3. Multiply the frequencies obtained in step 2 for each group by the weight (see Table X) to obtain the weighted food group score.
4. Sum the weighted food group scores, which provides the food consumption score for the households.
5. Use the appropriate thresholds to classify the households (see the unit of analysis section for metric 6)

Table 4. Food groups and weights for food consumption score calculation.

	Food Items	Food Groups (definitive)	Weights (Definitive)
1	Maize , maize porridge, rice, sorghum, millet pasta, bread, and other cereals Cassava, potatoes and sweet potatoes, other tubers, plantains	Main staples	2
2	Beans; peas, groundnuts, and cashew nuts	Pulses	3
3	Vegetables, leaves	Vegetables	1
4	Fruits	Fruit	1
5	Beef, goat, poultry, pork, eggs, and fish	Meat and fish	4
6	Milk, yogurt, and other diary	Milk	4
7	Sugar and sugar products, honey	Sugar	0.5
8	Oils, fats, and butter	Oil	0.5

9	Spices, tea, coffee, salt, fish power, small amounts of milk for tea.	Condiments	0
Source: WFP, 2008.			

Unit of analysis

The food consumption score normally groups the households' food security using the following thresholds for the score:

1) Poor, from 0-21,
2) Borderline, between 21.5 – 35
3) Acceptable for a score greater than 35.

A *poor* score may imply that the household is falling short of consuming at least one staple food or vegetable in the week, while the *acceptable* score is based on the consumption of oil and pulses in addition to staples and vegetables (Jones et al., 2013; WFP, 2008; Weismann et al., 2009).

Limitations

The measure is easy to administer and where standardized weights and cutoffs are used, the results may be compared across groups. But the cut off points should be examined depending on the context to ensure differentiation between household in that location. In areas where almost every household consumes a particular group such as sugar and oil, the cutoffs should be adjusted to ensure correct assessment of the performance of households in that area (Jones et al. 2013).

Nutritional status: Anthropometric measurements

Anthropometric measurements include height, weight, middle upper arm circumference[8], age, and gender. The measurements are done on children under the age of five years and women of childbearing age or mothers of children within the household being surveyed. The goal of the indicator is to assess prevalence of malnutrition, evaluate the impact of interventions on nutrition, enable the identification of at-risk persons, and monitor the nutritional status of a household or community over time (Jones et al., 2013). The measurements at the household level may be used, depending on sampling design, to calculate metrics at the community level to measure, for instance, rate stunting and wasting, percent of respondents underweight, and average birth weight (if the weight is collected or enumerated).

Method of data collection and data needed to compute metric

The data for the anthropometric measurements is collected as part of the household survey in most multi-indicator surveys. It is a separate section that requires specific equipment to ensure that the measurements are done correctly and at a time when the children and the mother of the children in the household (or a woman between the ages of 15 and 45) are present. The women are measured for indicators that determine status of wellbeing for women of reproductive age. To collect the data on height, weight, and mid upper arm circumference, data collectors need the following items: an adult digital weight scale, Leicester height measure, length mat to measure children under the age of 2 years, mid-upper arm circumference tape measure, and height stick about 100 centimeters long for children 2 to5years old. The weight should be measured in kilograms, height in centimeters, and age in months for children below five years (max 60

[8] Proxy for under nutrition

months). This will ensure that the data can be accurately analyzed by the software. The procedure and measurement are laid out in the Living Standards Measurement Study (LSMS) (World Bank, 2017) survey and the Vital Signs Protocol for household survey (VS, 2014)

Unit of analysis

Once the data on height, weight, mid upper arm circumference, the age and gender of the children and mother of children is recorded, the estimation are performed to determine children who are stunted, underweight, wasting, and overweight. The steps below indicate the algorithms and steps needed to construct the outcome indicators; % stunting, % wasting, % underweight, and % over weight. The World Health Organization has developed a software package that enables one to generate Z-scores for; weight-for-age (underweight), height-for-age (stunting), and weight-for-height (wasting) scores using the international growth standards for children under five years. The software can be downloaded from the WHO website (WHO, 2017). The scripts were developed for R, STATA, SAS, SPSS, and S-Plus software and you can choose the package with which you are most familiar with. An example is provided in the Appendix for STATA but for R can be found on this link and WHO site (http://www.who.int/childgrowth/software/en/). An analysis is then done for weight-for-age (zwei), height-for-age (zlen), and weight-for-height (zwfl) z-scores, and a body mass index (BMI) is calculated.

Underweight, stunting, overweight, and wasting

Percentage underweight is defined as the weight for age z-score (zwei) less than $(-)2\ sd$. Severe underweight is a z-score for weight for age such that $zwei < -3sd$. Extreme (i.e., biologically implausible) z-scores for each indicator are flagged. Stunting is defined as circumstance where a height for age z-score is more than two standard deviations below average. Severe stunting is defined as z-score for height-for-age <-3sd. Extreme (i.e., biologically implausible) z-scores for each indicator are flagged. Wasting is defined as the weight for height z score less than 2 standard deviations. Severe wasting is defined as z-score for weight-for-height <-3sd. Extreme (i.e., biologically implausible) z-scores for each indicator are flagged. Overweight is defined as BMI greater than 25. Obesity is defined as BMI >30. Extreme (i.e., biologically implausible) BMI scores are flagged according to the following system: BMI < 5 or zwfl > 60. Also note that BMI under 18.5 is underweight (Remans et al., 2015).

Limitations

Using anthropometric measurements to assess nutritional impact of an intervention are very important but have a few limitations. First, in order to have a robust estimate of these metrics, a large sample size is needed from the population surveyed and households with children under the age of five (60 months) should be sampled. This data collection can be costly to gather. Secondly, changes in nutritional outcomes have to be observed overtime, which means that the data collection must be done at baseline and for subsequent years. The target population, that is children and women, may also fluctuate across the time of study and this should be taken into consideration. For example, children who are four years old at the time of the baseline study, may exit the sample the next year of the survey since they will be over five years (60 months). These issues should be addressed with nutrition experts as sampling is done to collect these indicators.

Measurement method 2: Anthropometric measurements

Collection of data on nutritional outcomes of the population of focus for some projects may be too costly and hence the need for alternative approaches of assessing the nutritional status of the population in area of study. Additional data on anthropometry, or calculated nutritional indicators, may be obtained from secondary data to provide an indication on the nutritional status of the area at baseline. In addition, health center or national statistical offices may have health records or statistics. The following sources can be explored to provide an indication of the nutritional status of the study population: demographic health surveys (DHS, 2017); living standard measurement study surveys (World Bank, 2017); multiple indicator cluster surveys conducted by UNICEF (UNICEF, 2017); and local and national health center records.

Metric 7: Uptake of essential nutrients

Description of metric 7

Uptake of essential nutrients is determined by a number of factors that include, the consumption of nutrient-rich food, the preparation of the food, and the type of diet of the person consumes. This will affect the bioavailability of macronutrients and micronutrients. In populations that are deficient in essential nutrients, blood tests have been used to examine the levels of iron and other micronutrients. A researcher may use these tests to examine whether there has been an improvement in the nutritional status of the individual. One of the most common assays is a test for iron, where iron levels are examined across the population during the intervention.

Method of data collection and data needed to compute metric

The use of blood tests requires the presence of a trained health worker. The sample of persons to be tested for this measure will be developed further with consultations of health workers, nutritionists, and statistician.

Unit of analysis

Limitations

This metric is may be costly and will require some stringent efforts to gain approval from the institutional review boards (IRBs) and permission from the respective countries or agencies to allow this intrusive tests. This may be costly time wise and so should be accounted for during the planning phase. This procedure may be effective for large sample where there is potential to observe consumption and employ tests to see changes in micronutrient bioavailability.

Metric 9: Market or landscape supply of diverse food

Description of metric

A supply of diverse foods in an area may indicate demand of diverse food by the population or the supply of diverse food by farmers. Collection of data on availability of diverse food at the market may provide an indication on the “accessibility of diverse and nutritious foods.”

Method of data collection and data needed to compute metric

Data on the availability of foods at the market may be obtained from a “community survey” where data on the goods available at the market is collected. While standard community surveys are focused on enumerating the price of food items at the market, a modification of this survey may be done in collaboration with a nutritionist where data on various foods from different food groups is enumerated. These food groups may be similar to those used for calculation of dietary diversity or may be adjusted by the researcher as needed.

Unit of analysis

The unit of analysis will be a simple count of the number of food items that are available at the market by food group. This may be done by classifying the foods available in the same food categories as those used for the dietary diversity score computation. This survey is best performed twice, once during the hunger season right before the harvest and, if possible, again right after the harvest to see the variation in availability.

Food Security Indicator

Description of indicator

Measuring food security is an important indicator that has evolved in concept but is defined as a state in which "all people at all times have the both the physical and economic access to sufficient food to meet their dietary needs for a productive and healthy life" (USAID, 1992). Food security has evolved from food availability to examining nutritional capabilities of the food that is produced or accessible to a household (Burchi and De Muro, 2016). With the development of the Sustainable Development Goals, there has been a challenge to examine agriculture, nutrition, and food security in an integrated manner (UNSCN, 2015). The Food and Agricultural Organization has defined three main pillars of food security as food availability, food access, and food utilization.

Metric 1: Food production (availability) (field and farm)

Description of metric 1

At the field and farm level, this food security metric focuses on food that is made available due to the new intervention. The measure focuses on converting this output into calories as the measurement unit per area of land used to produce them. Data on production and area used for producing food items may be collected by survey or field measurement.

Method of data collection and data needed to compute metric

Following are the data needed to compute this metric:

1. Data on area used to produce the food item
2. Total amount of food that is produced on this area
3. Food composition tables that are available showing the calories in the food item (FAO, 2016)

The data on area and production can be obtained either from survey or field measurements (GPS for area measurement) or crop cuts (see productivity section on yields recall).

Unit of analysis

Following are the ways to calculate of calories of food item per hectare:

1. Obtain the mass of the food item produced and multiply it by the number of calories contained in 100 g of this food item.
2. Divide the total calories obtained by the area in hectares to obtain the metrics: calorie per hectare. Please note that area is at times reported in various units, e.g., acres, ropes, poles. Conversion units should be used to convert these into hectares.

Limitations

The measure at the field and farm level, if done on an experimental plot, may lack the data on household members (consumption aspect) and may not reflect the nonlinearity in consumption demands at the household. For example, a household of five adults will require more total calories than a household with two adults and three children.

Metric 2: Food availability (farm)

This measure is obtained by summing the total calories per food item (see food availability measure) and dividing them by the total land area used to produce these food items.

Metric 3: Food availability (household)

Description of metric

Food availability is defined as the availability of sufficient quantities of food of appropriate quality supplied through domestic production or importation (bought outside the household or scale or reference). Food availability indicator measures the amount of food produced by the household, and the amount that is sold and purchased per capita in order to estimate the calories and nutrients available per capita (Remans et al. 2013). In addition, this measure may also include a subjective food availability index that may be based on household report on the number of months or days of food insecurity. At the field level, scientists may use an alternative approach to estimate the potential calories available to the household at a given production level. This may be done by estimation of the total calories and nutrients produced and using participatory approaches to infer how the farmer might use this crop or livestock output.

Method of data collection and data needed to compute metric

Data to estimate this metrics are usually collected using a survey at the household level. The main data needed from the surveys for calculation include the following:

- Amount of crops produced by the household in standard metric units (e.g., kilograms, liters)
- Amount of crop produce that is sold
- Amount of food that is purchased from the market for consumption
- Amount of livestock and animal based products that is produced
- Amount of livestock and animal based products that is sold
- Household composition (e.g., members by age and gender)

The data on the household composition (age and gender of household members) are useful in determining the dietary intake of each individual with a focus on calories and nutrients. Food composition tables are used to provide an estimate of the amount of calories in a 100 g of a given crop (e.g., maize) and the nutrients contained in that crop. The food composition tables have been developed by the Food and Agricultural Organization (FAO) (FAO, 2016). Additionally, universities such as the Harvard University School of Public Health have developed food composition tables that are country specific, for example one developed for Tanzania (HSPH, 2016). The food composition tables provide data on the calories and nutrient per food item. For the analysis, the focus is on seven essential nutrients: carbohydrates, protein, vitamin A, vitamin C, iron, zinc, and folate. An adult equivalent scale is used to estimate total food requirements of an individual. Adult equivalent values provide an estimate by age and gender of caloric requirement of an individual based on the mean requirements by age and gender (Claro et al., 2010). In addition, dietary reference intake (DRI) values may be used in this estimation. The DRIs provide quantitative reference values of the nutrient intakes needed for a healthy living (USDA, 2017). For the estimation, we need to account for amount of food waster or lost postharvest. Data on food waste from production may be obtained either from the survey tool or from secondary sources such as FAO estimate of lost production from postharvest loss, consumption, and processing (FAO, 2011)

Unit of analysis

The food availability indicator is scored from 0 to 1. The calories or nutrients available per capita are estimated as follows:

1. To calculate the amount of food available per crop, subtract the amount sold and amount lost to waste from the total amount produced by the household.
2. To find the total amount available, add to this estimate the total amount of this crop that the household bought for consumption. The total food items available are then multiplied with the data from the food composition table to obtain the nutrients (seven nutrients) and calories (in kilo calories per 100 g). For example, if maize grain is one of the food items, obtain the number of calories per 100 g of maize grain from the food composition table. In addition, the table contains the nutrients content per item. The data from the seven nutrients should be listed. If the nutrient is not part of maize grain, indicate a "0"
3. To obtain the amount of calories and nutrient available per capita per day for the households, sum up the calories and nutrients across all the food items and divide by the adult equivalent of the household.

Metric 4: Food access

Description of metric

Food access may be defined as the ability to acquire sufficient quality and quantity of food to meet the nutritional requirements of individuals within the household for a productive life (Swindale and Bilisky, 2006). A food access indicator tends to focus on the economic aspect that examines the ability of household or person to purchase food. The indicator includes a computation of the household percentage of total expenditure allocate to food and minimum cost of a nutritious diet. These measures are used to examine if household income and expenditure may be able to achieve this minimum diet. This indicator is computed at the household level because the data on consumption expenditure and household composition (members in the household) are collected at the household scale.

Method of data collection and data needed to compute metric

Data to compute this metric is collected through a survey on the household composition, expenditure, and consumption of food items. The surveys used usually have a seven-day recall on the food consumption and for the nonfood consumption. The recall period for nonfood expenditures varies from one week to one year depending on the purchase and use of the item. For example, cigarettes or tobacco may have a one week (7-day) recall period; electricity bill payment at a one-month recall period; and building items such as cement, a 12-month recall period. The LSMS survey provides further details (World Bank, 2017).

Unit of analysis

The total household consumption measure from this section is similar to that used to compute the poverty metrics of household per capita expenditure. But the total expenditure for the poverty measure includes both the food expenditure and nonfood expenditure. For this metric, we further compute the minimum cost of local nutritious diet. This is used to compare the minimum cost of diet to the per capita expenditure of food. Use the following to compute the minimum cost of the diet:

1. Apply for access to software tool at http://www.savethechildren.org.uk/resources/online-library/the-cost-of-the-diet
2. Enter FAO local food price data – or data from household survey into the tool
3. Enter (average) household composition into the tool
4. Run tool; this provides the cost in $ for a minimum local nutritious diet for the (average) household per year
5. Divide by 365 days and the (average) number of people in the household to obtain the minimum cost of a nutritious diet per day per capita

From the nutritious diet, a gaps assessment is analyzed on the difference between the amount spent on food and the cost of a nutritious diet. A simple calculation of the buffer between what the household spends on food and total consumption expenditure is also made to come up with the food access score.

$$Gap = per\ capita\ food\ consumption\ /Minimum\ cost\ of\ a\ diet$$

$$Buffer = \left(1 - (food\ expenditure\ per\ capita/total\ expenditure\ per\ capita)\right)$$

$$Food\ access\ score = Gap * buffer$$

Metric 4: Food security – month of food insecurity

Description of metric

Month of food insecurity is an indicator used to assess the frequency of household food insecurity and the months in which these incidents occur. By asking a few questions, this method can be used as a quick measure of whether or not households face food insecurity. The food insecurity questions are found in most multi-indicator surveys such as the LSMS and are used as a subjective measure of food insecurity.

Method of data collection and data needed to compute metric

Data on the month in which the household did not have enough food (food insecure) are enumerated via survey at the household level. The following survey questions are asked of participants:

1. In the last 12 months, have you been faced with a situation when you did not have enough food to feed the household”
2. When did you experience that incident? (Months of the year are provided or listed and the household indicates which months.)

These questions are in the LSMS surveys (World Bank, 2017) and other multi-indicator surveys.

Unit of analysis and algorithm used for estimation

Computation of this data may require a few calculations and data staging. Ask the household member participant to indicate which months they did not have enough for consumption in the past year, Then code those months as one (1) or zero (0). Calculate to obtain the number of months that the household did not have enough food by summing up the response across the twelve months. The highest value that may be observed from this summation is 12. Divide the total months by 12. This will provide you a value from 0 to 1. This indicator can be used as a proxy measure for food insecurity.

Limitation

The limitation is that it will not provide you with information on what foods or food groups were not available that caused this insecurity. To obtain more robust information, other indicators listed previously may be used.

Metric 5: Rating of food security

Description of metric

Rating of food security is an indicator that is obtained using a participatory approach to examine farmer perceptions on the ability of the crop to provide enough food during the year.

Method of data collection and data needed to compute metric

This measure involves farmers rating the new technology compared to the conventional on its ability to provide enough food for consumption. Farmers may be asked, for example, if they prefer a new technology to a conventional one; which of two crops may provide more food for the household; and how they would rate the new technology in alleviating food insecurity (providing enough food throughout the year). The ratings are scaled as 1 for poor, 2 for intermediate, and 3for good. These ratings can be used to scale up across groups to assess the potential of the technology.

Food Safety

Description of indicator

Food safety is a key issue that ensures both a fit for consumption and quality of the food. Mycotoxins are often cited as potential issue that affects food safety that may lead to chronic illness when excessive mycotoxins exist in a given product (Milicevic et al., 2010). Mycotoxins often emerge due to certain weather conditions during the growing season or to poor storage. In addition, postharvest losses due to toxicity occur during agricultural production. These post-harvest losses may occur during harvesting, postharvest handling and storage, processing and distribution, and/oe during preparation and consumption. The most common measure of postharvest loss is through a questionnaire where the household is asked whether they incurred any postharvest loss; the cause of the loss, and what proportion of the harvest was lost (Kamiski and Christiaensen, 2014).

Pesticide contamination is an issue that also requires additional attention in food safety and quality. Pesticide contamination has been observed to be fatal in cases where quantities in food products exceed safe consumption levels. Unsafe levels can be detected through testing and enforced through regulation. In developed economies, these tests have been instituted but in developing economies where new technologies are disseminated or tested, measures to examine the contamination and its effect on human health is required.

Metric 1: Mycotoxin

Description of metric

Mycotoxins are toxic metabolites produced by fungus that occurs in food and feed commodities with the capability of causing disease and death among humans and animals. Human food is contaminated by mycotoxins from preharvest to postharvest (along the food chain). Humans are exposed to mycotoxin through consumption of contaminated plant-based foods or carryover of mycotoxins and their metabolites in animal derived products like meat and eggs (Milicevic et al., 2010; Boudra et al., 2007). Mycotoxins subject humans to chronic conditions that can be carcinogenic, mutagenic, teratogenic, estrogenic, estrogenic, hemorrhagic, immunotocxic, and neurotoxic (Milicevic et al., 2010). There are more than 300 known mycotoxins but focus has been on those that are toxic or carcinogenic. Aflatoxin B_1 is proven to be the most potent of these and also known to be genotoxic (can damage genetic information and lead to cell mutation that causes cancer). Studies indicate that Aflatoxin B_1 accounts for 75% of all contaminations of food plus feed products and is the most widespread among mycotoxin worldwide (Hussein and Brasel, 2001).

Method of data collection and data needed to compute metric

The testing of contamination of the food products requires a sample of the food crops or animal product at a given stage of harvest or storage. For aflatoxins, the gold standard measures are the Thin-layer Chromatography (TLC) (de Iongh et al., 1964) and the High Performance Liquid Chromatography (HLPC). The later was developed as an improvement to TLC to ensure an efficient and automated process. These methods are widely used but are expensive and require a long time for preparation of the samples. Therefore, an alternative measure was developed, Immunoassay (Berson et al., 1968). Immunoassay is now used for onsite detention of mycotoxins in food and animal feed (Wacoo et al., 2010). Currently, Enzyme-Linked Immunosorbent Assay (ELISA) (Avrameas, 1969) is widely used in detection of aflatoxin in agricultural food products (Wacoo et al., 2010; Devi et al., 1999).

ELISA-test toolkits, based on immunoassay, are now widely used for onsite tests (Huybrechts, 2011) for mycotoxins for several reasons: 1) the tool kits are inexpensive and easy to use; 2) a large number of samples can be analyzed at the same time; and 3) compared to earlier radio immunoassay, they reduce the health hazards associated with test materials (Twyman, 2005; Wacoo et al. 2010). In recent usage scientists at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) indicated that the use of tool kits from the ELISA family that is, competitive enzyme-linked immunosorbent assay (CELISA), reduced the cost from $25 to $1 per sample (ICRISAT, 2009).

Unit of analysis

The unit of analysis for this indicator will be toxicity units per gram of agricultural food product.

Limitations

ELISA test kits, although recommended, may still be costly. Therefore, a project needs to ensure that the benefits for testing for mycotoxins are assessed to justify the need for undertaking these measurements. In addition, the use of this technique requires skilled technicians and safety measures to ensure that areas are sanitized and cleaned well after testing.

Metric 2: – Pesticide contamination

Description of metric

Pesticides cover a range of compounds that include insecticides, fungicides, herbicides, rodenticides, molluscides, nematicides, plant growth regulators among others (Aktar et al., 2009). Indicators focus on the risk and environmental impact of pesticides on water quality, human health, and death of species (Padovani et al., 2004; Damalas and Eleftherohorinos, 2011). Pesticides have benefits that include improving productivity through protection from crop losses, vector control that reduces diseases, and indirect impacts like increased food security and nutrition. On the cost side, pesticides are dangerous to the targeted and nontargeted species and when used in excess amounts may kill nontargeted plants, insects, and animal species, which can affect biodiversity and other ecosystems services (Padovani et al., 2004). In addition, pesticides may contaminate agricultural foods and exposure at application may also have effects on human health (Dagupta et al., 2007; London and Bailie, 2001; Aktar et al., 2009; Damalas and Eleftherohorinos, 2011).

The first mass cases of pesticide poisoning were reported in India in 1958 where 100 people died after consuming contaminated wheat flour (Karunakaran, 1958). This raised awareness for the need to test and examine the levels of pesticide contaminations in food in the years that followed in India. In Europe, programs began examining the levels of pesticides in food products as early as 1999, when the European Union commissioned "Monitoring of Pesticide Residues in Products of Plant Origin in the European Union, Norway, and Iceland 1999 Report" (SANCO, 2001). Such reports have continued.

The data on risk of exposure to pesticide by farm workers is scanty yet high levels of exposure have been reported above the allowable limits in some studies (Dasgupta et al, 2007). For example, a study in Vietnam that took blood samples from rice farmworkers who had been exposed to insecticides (e.g., compounds containing carbamates and organophosphates) associated with health risks such as DNA damage, birth defects, hormonal changes and fetal death (Zahm et al, 1997; Dasgupta et al, 2007). Thirty-five percent of the farmers (n=190) were reported to have acute pesticide poisoning. However, it should be noted that blood tests are not always performed to assess the levels of poisoning in farmworkers. Instead subjective surveys are performed to assess symptoms and these may not provide conclusive results (Dasgupta et al., 2007). Subjective surveys and blood and health tests should be performed in tandem. Increased surveillance of pesticide poisoning and contamination in agricultural foods is also needed to understand the potential effect and how these levels are associated with some demographic and socio-economic factors (Aktar et al., 2009). In this section we cover both aspects of measurement such as the chemical tests and health center data.

Method of data collection and data needed to compute metric

Pesticide contamination indicators and metrics need to be integrated in projects where pesticide use is an intervention. But measurement of the effects of these indicators and data collection may require expert knowledge of the following: 1) the composition of the pesticide being used and 2) the effects of those components on human health (symptoms of exposure). Having this information may allow those gathering health data to examine whether some symptoms that are reported in those areas may be linked to pesticide use. Additionally, researchers can determine more direct measurement is required for the exposed population.

The most direct method is collection of data from the exposed population. This should be done with the help of a health worker. A selected sample should be tested for exposure to the given chemical contaminants from the pesticides. Studies such as Dasgutpa et al. (2011) have used blood samples from the populations that were subsequently analyzed in the lab for the presence of contaminant in blood.

Unit of analysis

The tests by health workers determine if the level of contaminant is above the maximum allowable level in humans. The units for assessment are obtained from the health specialist when needed.

Limitations

The main limitations to pesticide assessment are the costs of assessments and sometimes a low level of response from policy makers.. Due to nonspecific information on the long-term effects of exposure to pesticides and the low levels of reported poisoning (compared to added benefits of pesticides), some areas have not invested in this research. Additionally, a thorough examination of the pesticide contaminations may be costly and require a high level of expertise.

Metric 2: Postharvest losses

Description of metric 2

Postharvest (PHL) loss reduces the amount of food that is available to households, which can affect household food security. Postharvest loss may occur at various stages: 1) during harvest from mechanical damage and spillage; 2) during postharvest handling, including damage in storage (rodents, pests, rotting) and during drying and winnowing; 3) during processing; 4) through distribution and marketing; and 5) during consumption (FAO, 2011; Kaminski and Christiaensen, 2014). With advances in technologies that are affordable, reducing postharvest loss is a growing initiative to increase returns on investment for farmers (Zorya et al., 2011). There has been a debate on how to measure PHL to ensure that there is comparison of loss across different crops and how loss may vary at different stages in the product life cycle. Some studies have used loss of food as expressed in calories (Lipinski et al., 2013) and other studies express loss in terms of percent of weight (FAO, 2011). This may be an analytical aspect to consider in analysis. Survey tools have in the past focused on postharvest loss at harvest and during storage for a particular season. At the household level in agricultural surveys, less emphasis is made on enumerating losses in processing food and during the marketing process, but there is an interest in loss of food while crops are in the field (Kaminski and Christiaensen, 2014). The latter has been enumerated looking at whether there was loss of crops as they were awaiting harvest but there is no emphasis on the quantity lost since it would be difficult for the farmers to estimate.

Method of data collection and data needed to compute metric

Data to estimate this metric are collected using a survey. In standard survey tools, the focus is too discover if there was a loss, what were the causes, and how much was lost. Following are sampe questions (NBS, 2014):

1. Was any portion of production lost postharvest?
2. What was the reason for the loss (rotting, insects, pests, theft, other)?
3. Out of 10 units, how many were lost?
4. What was the value of the lost crop?

Unit of analysis

The unit of analysis is the percentage of harvest that was lost. This can be obtained from the question 3.It is important to perform quality checks of the data, for example, to remove data errors as well as typographical errors (some enumerators may enter percentages instead of unit from 1 to 10).

Limitations

In the new versions of surveys, data are enumerated on preharvest losses in the field and not much is enumerated on losses due to processing and marketing. There may be an over estimation of postharvest losses, depending on definition, if the farmer assumes that there is need to include all the losses incurred from harvest to marketing and the researcher is only focused on premarketing losses (Kaminski and Christiaensen, 2014). This issue may be addressed during enumerator training. Additionally, farmers that incurred small losses may not be inclined to report them and this may underestimate the PHL for the population.

Human Health

Description of indicator

Human health may be at risk due interaction with animals or through vector borne diseases that affect animals and humans. For example, chicken production in close proximity to the household may increase the risk of Newcastle disease virus; which can be transmitted to humans. There are also other livestock diseases that can be transferred to humans in addition to other contaminants from fecal matter. Campylobacter bacteria from animal feces infect humans through contaminated food and may attack the humans small and large intestines, which causes chronic illness. In some cases, this contamination has been associated with stunting in children.

Metric 1: Incidence of zoonotic diseases

Description of metric

Zoonotic diseases are diseases or infections that can be transferred directly or indirectly between humans and animal through consumption of contaminated animal products or byproducts or through contact with animals. There are a number of bacteria that can cause food poisoning, including *Campylobacter spp*, *E.coli*, and *Salmonella spp*. as well as viruses that can cause sickness (e.g., viruses avian and swine influenza). *Campylobacter spp*, *E.coli*, and *Salmonella spp*. are bacteria that cause food poisoning following consumption of contaminated livestock products. For example, *Enterohaemorrhagic E.coli* (EHEC) is a common strain carried by cattle, sheep, and pigs that through fecal material can contaminate meat and animal products. This *E.coli* species may detected in the intestines or hides of catch at slaughter or in meat samples at processing (Elder et al., 2000; Waller et al., 2010). Studies in the United States and Europe have found EHEC in 20-28% of hides and intestines, and 43% of meat samples at processing (WSPA, 2013; Waller et al., 2010; Elder et al., 2000). Although infections from this pathogen may be rare, the fact that they may be fatal in nature makes it a critical to examine and prevent (Pennington, 2010). Campylobacter is one of the largest cause of incidences of gastroenteritis (a stomach or intestinal infection that causes diarrhea, nausea, vomiting, and fever) (WSPA, 2013; WHO, 2012). Poultry is a major source of campylobacter and may be in the meat, liver, and intestines of poultry. However, it can also reside on the outside (skin) (EFSA, 2010) of other farm animals. Campylobacter is also a serious issue in the European Union where tests have found in some instances infections in more than 75% of the chicken on sale (EFSA, 2016; WSPA).

In the developing world, due to low levels of surveillance, conclusive evidence is lacking on the actual prevalence of campylobacter, but data from clinical results indicate a high prevalence (Platts, Mills and Kosek, 2014; Lengerh et al., 2013). In developing countries, the risk factors are associated with environmental hazards such as drinking contaminated water. Additionally, incidences of campylobacter are more common in children than in adults (Coker et al., 2002). Campylobacter incidences have been observed with high incidences with ability to be associated with diarrhea in Malawi and temporary reduction in weight gain in children (3 months) (Lee et al., 2013). Studies also indicate that campylobacter in poultry may be reduced through the following measures: reducing acute stress of the birds during the process of transportation; monitoring practices such as “thinning” and “fasting” that increase bird stress, which predisposes birds to bacterial infection (Northcutt et al, 2003; WSPA, 2013)

Method of data collection and data needed to compute metric

Data on prevalence of zoonotic diseases in these areas may be obtained from health centers. The cost of obtaining stool and blood sample to perform these tests on the human population may be cost prohibitive. If researcher knows the symptoms associated with a zoonotic disease and the pathogens, then data may be collected from the individuals via a survey, or the researcher may ask health centers if most of the population tend to show such symptoms.

Unit of analysis

The unit of analysis is the number of persons infected by the disease per 100,000 of population.

Limitations

The limitation of this indicator and data source is that if there are no functional health centers where such data is collected, then assessing prevalence may be difficult and the project must decide whether or not to invest in such data collection for zoonotic disease prevalence.

Metric 2: Incidence of vector-borne diseases

Description of metric

Vector-borne diseases are those infections that are transmitted between human or from humans to animals by vectors such as arthropod species (e.g., mosquitoes, ticks, black flies) (WHO, 2016). Mosquitoes are the most widely known disease vectors, and they that cause malaria, dengue fever, Rift Valley fever, yellow fever, West Nile fever, Japanese encephalitis, and lymphatic filariasis (Who, 2014). The link between vector-borne diseases and agricultural production imposes burdens on the population that affect productivity as well as animal and human health.

Method of data collection and data needed to compute metric

Although there are studies that have assessed burden of disease using Disability Adjusted Life Years (DALY[9]) (Murray and Lopez, 1996) as a metric to examine the economic burden of the disease for the population, in this metric we focus on the incidence of the disease. This Measurement of incidence of the disease among a given population may be costly due to need to test individuals using human samples (blood or stool). For this metric, we recommend using available data from health centers in the location where the project is taking place. The researcher should identify the health center that is close to the areas where of data collection. These centers may have data on the reported cases and can provide a general understanding on health status of the various populations within the area.

Unit of analysis

The unit of analysis of incidences of a disease are mainly provided at cases per 1,000 of population.

Limitations

This measure would be limited by the lack of health centers close to a study area. Another limitation may be that the data collected in that area may not be available to be shared to the public. If this occurs, the project researchers may look at alternative options, such as collecting data over the life of the project (if the net benefits of such information to the project significantly outweigh the costs).

[9]The DALYs are beyond the scope of this assessment

Capacity to experiment

Description of Indicator

Capacity to experiment is ability of the household to use or adopt (experiment with) a new technology. This may be assessed by noting the number of new technology components that a household or farmer is able to adopt. A scale should be created to score a range of abilities to adopt new technology from none to proficient.

Metric 1: Number of new practices being tested

Description of metric 1

The capacity to experiment metric focuses on the number of technological practices or components that a farmer is able to test. The data collection may be done via survey or focus group discussions. In a survey, farmers are asked about the management practices that he/she applies on the field. This should include the technologies that the researcher would like to assess. For example, if the technology is assessing Integrated Soil Fertility Management (ISFM), all the components of ISFM used by the farmer should be counted, including for example, improved germplasm, chemical fertilizer, organic fertilizer use and management, and adaptation to local conditions (e.g., erosion control, use of lime on acidic soils) (Vanlauwe et al., 2012).

Method of data collection and data needed to compute metric

To assess this measure, farmers are asked during survey or focus group to indicate what aspects of a technology or intervention they are currently using. A list may be provided for them to choose the relevant modalities. For example, in the ISFM situation, a farmer would indicate which or what parts of the three choices are used. In case technology packages, a similar approach should be taken.

Unit of analysis

The unit of analysis is a count of the number of practices that are implemented by the farmer. The maximum number of practices is the total number of practices that the farmers is exposed to during the interventions period (3 in ISFM example).

Limitations

This indicator is a simple and straight forward way to assess capacity to experiment, but it may not answer the "why" questions. Therefore, researchers may complement this indicator with other indicators within the framework.

Metric 2: Percentage of farmers experimenting

Description of indicator

This indicator is similar to the previous one and examines the number of farmers participating in a given technology of the total farmers exposed to it.

Method of data collection and data needed to compute metric

This approach may be done by survey of representative sample of the farmers in the area. As in the indicator "Number of new practices being tested," a similar data collection approach is taken to obtain information on the farmers who may be experimenting with a given technology. Questions such as the following can be asked: On your farm, do you irrigate crops during the season? If so, what type of irrigation technique do you use? Such questions, for example, may be used to identify which households are using irrigation technology that is new in that area. Other questions would need to be developed for other technologies.

Unit of analysis, and algorithm used for estimation

The unit of analysis is the percentage of household experimenting with a technology. A summation of the number of the households that indicated using a technology it obtained and divided by the total number of households in the area.

References

Aktar, W., Sengupta, D. and Chowdhury, A., 2009. Impact of pesticides use in agriculture: their benefits and hazards. Interdisciplinary toxicology, 2(1), pp.1-12.

Allen, L.H., 2000. Ending hidden hunger: the history of micronutrient deficiency control. Background Analysis for the World Bank-UNICEF Nutrition Assessment Project World Bank.

Anderson, J. and Ingram, J. 1993.Tropical Soil Biology and Fertility – A Methods Manual 2nd edition, CABI.

Andow, D.A. 1991. Vegetational diversity and arthropod population response. Annu. Review Entomol. 36:561-586.

Approaches for Rural and Urban Areas http://dhsprogram.com/publications/publication-wp60-working-papers.cfm

Bellon, M.R., 2001. Participatory research methods for technology evaluation: A manual for scientists working with farmers. CIMMYT.Blackwell Scientific Publications, Oxford.

Berson, S.A. and Yalow, R.S., 1968. General principles of radioimmunoassay. Clinica chimica acta, 22(1), pp.51-69.

Boudra, H., Barnouin, J., Dragacci, S. and Morgavi, D.P., 2007. Aflatoxin M 1 and ochratoxin A in raw bulk milk from french dairy herds. Journal of dairy science, 90(7), pp.3197-3201.

Burchi, F., Fanzo, J. and Frison, E., 2011. The role of food and nutrition system approaches in tackling hidden hunger. International journal of environmental research and public health, 8(2), pp.358-373.

Carter, M. R., & Barrett, C. B. (2006). The economics of poverty traps and persistent poverty: An asset-based approach. The Journal of Development Studies, 42(2), 178-199.

Claro, R.M., Levy, R.B., Bandoni, D.H., Mondini, L., 2010. Per capita versus adult-equivalent estimates of calorie availability in household budget surveys. Cadernos de Saúde Pública, 26(11), pp.2188-2195.

Coe, R. 2002. Analyzing Ranking and Rating Data from Participatory On-Farm Trials. p.44 in Bellon, M.R. and Reeves, J., *Quantitative analysis of data from participatory methods in plant breeding*. CIMMYT. http://libcatalog.cimmyt.org/download/cim/448231.pdf

Coker, A.O., Isokpehi, R.D., Thomas, B.N., Amisu, K.O. and Obi, C.L., 2002. Human campylobacteriosis in developing countriessynopsis-statistical data included. Emerg Infect Dis, 8(3), pp.237-243.

Damalas, C.A. and Eleftherohorinos, I.G., 2011. Pesticide exposure, safety issues, and risk assessment indicators. International journal of environmental research and public health, 8(5), pp.1402-1419.

Dasgupta, S., Meisner, C., Wheeler, D., Xuyen, K. and Lam, N.T., 2007. Pesticide poisoning of farm workers–implications of blood test results from Vietnam. International journal of hygiene and environmental health, 210(2), pp.121-132.

De Groote, H., Rutto, E., Odhiambo, G., Kanampiu, F., Khan, Z., Coe, R. and Vanlauwe, B., 2010. Participatory evaluation of integrated pest and soil fertility management options using ordered categorical data analysis. Agricultural Systems, 103(5), pp.233-244.

De Iongh, H., Vles, R.O. and De Vogel, P., 1964. The occurrence and detection of aflatoxin in food. In Symposium on Mycotoxins in Foodstuffs ed. GH Wogan, MIT Press, Cambridge, Massachusetts. Google Scholar (p. 235).

de Pee and Bloem, W., 2007. The bioavailability of (pro) vitamin A carotenoids and maximizing the contribution of homestead food production to combating vitamin A deficiency. International Journal for Vitamin and Nutrition Research, 77(3), pp.182-192.

Deaton, Angus, and Salman Zaidi. Guidelines for constructing consumption aggregates for welfare analysis. Vol. 135. World Bank Publications, 2002.

Deaton, Angus. The analysis of household surveys: a micro-econometric approach to development policy. World Bank Publications, 1997.

Demographic Health Survey. (DHS) 2008. The DHS Wealth Index:

Demographic Health Survey. 2017. Data. https://dhsprogram.com/Data/ Accessed January 2017.

Devi, K.T., Mayo, M.A., Reddy, K.L.N., Delfosse, P., Reddy, G., Reddy, S.V. and Reddy, D.V.R., 1999. Production and characterization of monoclonal antibodies for aflatoxin B1. Letters in applied microbiology, 29(5), pp.284-288.

EFSA Panel on Biological Hazards (2011) Scientific Opinion on Campylobacter in broiler meat production: control options and performance objectives and/or targets at different stages of the food chain. EFSA Journal, 9(4): 2105. http://www.adiveter.com/ftp_public/A4150411.pdf

Elder, R.O., Keen, J.E., Siragusa, G.R., Barkocy-Gallagher, G.A., Koohmaraie, M. and Laegreid, W.W., 2000. Correlation of enterohemorrhagic Escherichia coli O157 prevalence in feces, hides, and carcasses of beef cattle during processing. Proceedings of the National Academy of Sciences, 97(7), pp.2999-3003.

FAO (2016) International food composition table/database directory. http://www.fao.org/infoods/infoods/tables-and-databases/en/

FAO. 2011. Global food losses and food waste – Extent, causes and prevention. Rome http://www.fao.org/docrep/014/mb060e/mb060e.pdf

FAO. 2013. Guidelines for measuring household and individual dietary diversity. http://www.fao.org/3/a-i1983e.pdf Accessed January, 2017

Fermont, A. and Benson, T., 2011. Estimating yield of food crops grown by smallholder farmers. *International Food Policy Research Institute, Washington DC*, pp.1-68.

Filmer, D., and Pritchett, L. H. (2001). Estimating wealth effects without expenditure Data—Or tears: An application to educational enrollments in states of india. Demography, 38(1), 115-132.

Food and Agricultural Organization (FAO) 2011. Global Food Losses and Food Waste-Extent, Causes and Prevention. FAO, Rome, Italy. http://www.fao.org/docrep/014/mb060e/mb060e.pdf

Food and Agricultural Organization. (FAO) 2017. FAO/INFOODS Food Composition Databases. http://www.fao.org/infoods/infoods/tables-and-databases/faoinfoods-databases/en/

Foster, J., Greer, J. and Thorbecke, E., 1984. A class of decomposable poverty measures. Econometrica: Journal of the Econometric Society, pp.761-766.

Ghana Statistical Service (GSS) 2016. Ghana - Population and Housing Census 2010 - IPUMS Subset http://microdata.worldbank.org/index.php/catalog/2151/study-description Accessed January, 2017

Harvard University's School of Public Health (HSPH) 2016. Tanzania Food Composition Tables. https://www.hsph.harvard.edu/nutritionsource/food-tables/

Hussein, H.S. and Brasel, J.M., 2001. Toxicity, metabolism, and impact of mycotoxins on humans and animals. Toxicology, 167(2), pp.101-134.

Huybrechts, B., 2011. Evaluation of Immunoassay Kits for Aflatoxin Determination in Corn & Rice. National Reference Laboratory on Mycotoxins. http://www.favv-afsca.fgov.be/laboratoires/laboratoiresagrees/_documents/2012-05-04_Evaluation_immunoassay_kits_aflatoxin.pdf

International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) (2009). Aflotoxin testing kit. Protects human health; helps meet international market standards. http://www.icrisat.org/impacts/impact-stories/icrisat-is-aflotoxin-kit.pdf

International Food Policy Research Institute. 2014. Global Nutrition Report 2014: Actions and Accountability to Accelerate the World's Progress on Nutrition. Washington, DC.

Jones, A.D., Ngure, F.M., Pelto, G. and Young, S.L., 2013. What are we assessing when we measure food security? A compendium and review of current metrics. Advances in Nutrition: An International Review Journal, 4(5), pp.481-505.

Jones, A.D., Ngure, F.M., Pelto, G. and Young, S.L., 2013. What are we assessing when we measure food security? A compendium and review of current metrics. Advances in Nutrition: An International Review Journal, 4(5), pp.481-505.

Jones, A.D., Ngure, F.M., Pelto, G. and Young, S.L., 2013. What are we assessing when we measure food security? A compendium and review of current metrics. Advances in Nutrition: An International Review Journal, 4(5), pp.481-505.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., 2003. The DSSAT cropping system model. European Journal of Agronomy 18, 235–265.

Kaminski, J. and Christiaensen, L., 2014. Post-harvest loss in sub-Saharan Africa—what do farmers say?. Global Food Security, 3(3), pp.149-158.

Karunakaran, C.O., 1958. The Kerala food poisoning. Journal of the Indian Medical Association, 31(5), pp.204-7.

Kawano, K., 1990. Harvest index and evaluation of major food crop cultivars in the tropics. *Euphytica*, 46(3), pp.195-202.

Kennedy, G.L., Pedro, M.R., Seghieri, C., Nantel, G. and Brouwer, I., 2007. Dietary diversity score is a useful indicator of micronutrient intake in non-breast-feeding Filipino children. The Journal of nutrition, 137(2), pp.472-477.

Lee, G., Pan, W., Yori, P.P., Olortegui, M.P., Tilley, D., Gregory, M., Oberhelman, R., Burga, R., Chavez, C.B. and Kosek, M., 2013. Symptomatic and asymptomatic Campylobacter infections associated with reduced growth in Peruvian children. PLoS Negl Trop Dis, 7(1), p.e2036.

Lengerh, A., Moges, F., Unakal, C. and Anagaw, B., 2013. Prevalence, associated risk factors and antimicrobial susceptibility pattern of Campylobacter species among under five diarrheic children at Gondar University Hospital, Northwest Ethiopia. BMC pediatrics, 13(1), p.82.

Lobell, D. B. The use of satellite data for crop yield gap analysis. *F. Crop. Res.* **143**, 56–64 (2013).
Marten, G. C, J. S. Shenk, and F. E. Barton II, editors. 1989. *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality.* U.S. Department of Agriculture, Agriculture Handbook No. 643 (revised with supplements), 110 p.

London, L. and Bailie, R., 2001. Challenges for improving surveillance for pesticide poisoning: policy implications for developing countries. International Journal of Epidemiology, 30(3), pp.564-570.

McKenzie, D. J. 2005. Measuring inequality with asset indicators. Journal of Population Economics, 18(2), 229-260.

Mead, R. and Willey, R.W., 1980. The concept of a 'land equivalent ratio' and advantages in yields from intercropping. *Experimental Agriculture*, 16(03), pp.217-228.

Michelson, H., M. Muñiz, and K. DeRosa. 2013. Measuring Socio-economic Status in the Millennium Villages: The Role of Asset Index Choice. The Journal of Development Studies, 49(7), 917-935.

Micronutrient Initiative (MI), 2011. Micronutrient Initiative Annual Report 2010-2011.

Milićević, D.R., Škrinjar, M. and Baltić, T., 2010. Real and perceived risks for mycotoxin contamination in foods and feeds: challenges for food safety control. Toxins, 2(4), pp.572-592.

Mourice, S. K. et al. 2014. Maize cultivar specific parameters for decision support system for agrotechnology transfer (DSSAT) application in Tanzania. *American Journal of Plant Sciences* 5.6: 821.

Muthayya, S., Rah, J.H., Sugimoto, J.D., Roos, F.F., Kraemer, K. and Black, R.E., 2013. The global hidden hunger indices and maps: an advocacy tool for action. PLoS One, 8(6), p.e67860.

National Bureau of Statistics (NBS) – Government of Tanzania. 2014. Tanzania National Panel Survey 2012-2013, Third Round. Dar es Salaam, Tanzania. NBS. http://www.nbs.go.tz/tnada/index.php/catalog/29

Newbould, P. J. (1967) Methods of Estimating the Primary Production of Forests.

Newton, A.C. G.S. Begg and J.S. Swanston. 2008. Deployment of diversity for enhanced crop function Annuals of Appl Biology 1-14

Padovani, L., Trevisan, M. and Capri, E., 2004. A calculation procedure to assess potential environmental risk of pesticides at the farm level. Ecological Indicators, 4(2), pp.111-123.

Palm, C.A., Gachengo, C.N., Delve, R.J., Cadisch, G. and Giller, K.E., 2001. Organic inputs for soil fertility management in tropical agroecosystems: application of an organic resource database. *Agriculture, Ecosystems & Environment*, 83(1), pp.27-42.

Peng, D., Huang, J., Li., C., Liu, L., Huang., W., Wang, F. & Yang., X. Modelling paddy rice yield using MODIS data. *Agric. For. Meteorol.* **184**, 107–116 (2014).

Pennington, H., 2010. Escherichia coli O157. The Lancet, 376(9750), pp.1428-1435.

Proctor, J. (1983) Tropical Forest Litterfall 1. Problems of data comparison. In: Sutton, S.L., Whitmore, T.C. and Chardwick, A.C. (eds.), Tropical Rain Forest: Ecology and Management. Blackwell Scientific Publications, Oxford, pp.267-273

Reeves, M. C., Zhao, M. & Running, S. W. Usefulness and limits of MODIS GPP for estimating wheat yield. *Int. J. Remote Sens.* **26,** 1403–1421 (2005).

Remans, R., DeRosa, K., Palm, C. 2015. Nutrition Thread. *Vital Signs Project Unpublished*.

Rossel, R.A.V. and Bouma, J., 2016. Soil sensing: A new paradigm for agriculture. *Agricultural Systems*, *148*, pp.71-74.

Rutstein, S. O. 2008. The DHS Wealth Index: Approaches for Rural and Urban Areas. Calverton, Maryland, USA: Macro International - See more at: http://dhsprogram.com/publications/publication-wp60-working-papers.cfm#sthash.dhPElqZu.dpuf

Rutstein, S.O., Johnson, K. and Measure , O.M., 2004. The DHS wealth index. ORC Macro, MEASURE DHS.

http://www.dhsprogram.com/programming/wealth%20index/Steps_to_constructing_the_new_DHS_Wealth_Index.pdf

Sachs, J. and Malaney, P., 2002. The economic and social burden of malaria. Nature, 415(6872), pp.680-685.

SANCO, 2001. Monitoring of Pesticide Residues in Products of Plant Origin in the European Union, Norway, and Iceland 1999 Report. http://ec.europa.eu/food/fvo/specialreports/pesticide_residues/annex_1_1999_en.pdf

Sapkota, T.B., Jat, M.L., Jat, R.K., Kapoor, P. and Stirling, C., 2016. Yield Estimation of Food and Non-Food Crops in Smallholder Production Systems. In: Rosenstock, T.S., Rufino, M.C., Butterbach-Bahl, K., Wollenberg, E. and Richards, M., (Eds.) *Methods for Measuring Greenhouse Gas Balances and Evaluating Mitigation Options in Smallholder Agriculture.* Springer

Shaxson, L. and L.W. Tauer. 1992. Intercropping and diversity: An economic analysis of cropping patterns on smallholder farms in Malawi. Experimental Agriculture 28:211-228.

Sibhatu, K.T., Krishna, V.V. and Qaim, M., 2015. Production diversity and dietary diversity in smallholder farm households. Proceedings of the National Academy of Sciences, 112(34), pp.10657-10662.

Stuth, J., Jama, A., Tolleson, D. 2003. Direct and indirect means of predicting forage quality through near infrared reflectance spectroscopy. *Field Crops Research* 84, 45–56

Swindale, A. & Bilinsky, P., 2006. Household Dietary Diversity Score (HDDS) for measurement of household food access: Indicator guide, Washington,

Twyman, R.M., 2005. Immunoassays, applications: Clinical. Worsfold P., Townshend A, Poole C. Encyclopedia of Analytical Science (2nd edn).–Elsevier Science, London, 4, pp.317-324.

U.S. Department of Agriculture (USDA) (2017) Dietary Reference Intakes. Available: https://www.nal.usda.gov/fnic/dietary-reference-intakes Accessed January 2017

UNICEF, 2017. Multiple Indicator Cluster Survey. http://mics.unicef.org/ Accessed January, 2017.

UNSCN. 2016. Investments for Healthy Food Systems: A Framework Analysis and Review of Evidence on Food System Investments for Improving Nutrition. Implementing the Framework for Action of the Second International Conference on Nutrition. Rome.

Vanlauwe B, Pypers P, Birachi EA, Nyagaya M, Van Schagen B, Huising J, Ouma E, Blomme G, Van Asten P (2012) Integrated soil fertility management in Central Africa: experiences of the consortium for improving agriculture-based livelihoods in Central Africa (CIALCA). In: Hershey CH (ed) Eco-Efficiency: from vision to reality. Centro Internacional de Agricultura Tropical (CIAT), Cali, CO. 17 p

Vital Signs (VS) 2014. Vital Signs Protocol. Household Survey. http://vitalsigns.org/sites/default/files/VS_Protocol_HouseholdSurvey_August2016.pdf

Wacoo, A.P., Wendiro, D., Vuzi, P.C. and Hawumba, J.F., 2014. Methods for detection of aflatoxins in agricultural food crops. Journal of Applied Chemistry, 2014.

Walker, C., Shi, X., Sanderson, M., Sargeant, J. and Nagaraja, T.G., 2010. Prevalence of Escherichia coli O157: H7 in gut contents of beef cattle at slaughter. Foodborne pathogens and disease, 7(3), pp.249-255.

Wiesmann, D., Bassett, L., Benson, T. and Hoddinott, J., 2009. Validation of the World Food Programme s Food Consumption Score and Alternative Indicators of Household Food Security. Intl Food Policy Res Inst.

World Bank 2017. Living Standard Measurement Study. http://go.worldbank.org/UK1ETMHBN0

World Bank 2017. Living Standards Measurement Study (LSMS) survey. http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,contentMDK:21610833~pagePK:64168427~piPK:64168435~theSitePK:3358997,00.html

World Food Program (WFP) 2008. Food Consumption Score – Calculation and use of the food consumption score in food security analysis. http://documents.wfp.org/stellent/groups/public/documents/manual_guide_proced/wfp197216.pdf

World Food Program (WFP) 2015. Food Consumption Score Nutritional Analysis (FCS-N) Guidelines. https://www.wfp.org/content/food-consumption-score-nutritional-quality-analysis-fcs-n-technical-guidance-note

World Health Organization (WHO) 2016. Vector Borne Diseases http://www.who.int/mediacentre/factsheets/fs387/en/

World Health Organization (WHO) 2017. WHO Anthro (version 3.2.2, January 2011) and macros. http://www.who.int/childgrowth/software/en/ Accessed January 2017.

World Society for the Protection of Animals (WSPA) 2013. Zoonotic diseases, human health and farm animal welfare. https://www.ciwf.org.uk/media/3756123/Zoonotic-diseases-human-health-and-farm-animal-welfare-16-page-report.pdf

Zorya, S., Morgan, N., Diaz Rios, L., Hodges, R., Bennett, B., Stathers, T., Mwebaze, P. and Lamb, J., 2011. Missing food: the case of postharvest grain losses in sub-Saharan Africa.

Social Domain: (Note: The superscript letters (a,b,c) after each metric refer to the methods in the right-hand column)

SOCIAL DOMAIN			
Indicator	**Field, farm, and household level metrics**	**Community/Landscape + metrics**	**Measurement method**
Gender equity	Resources: Land access by gender [a-d] Livestock ownership by gender [a-d] Capacity: Access to information [a-d] Agency: Time allocation by gender [a-d] Management control by gender [a-d] Market participation by gender [a-d] Achievements: Income by gender [a-d] Nutrition/Food security by gender [a-d] Health status by gender [a-d] Cross-cutting: Rating of technologies by gender [b]	Women Empowerment in Agriculture Index [a, d] (measures absolute and relative empowerment across five domains: production, resources, income, leadership, and time)	[a] Individual survey [b] Participatory evaluation [c] Focus group discussions [d] Household survey
Equity (generally)	Access to resources (land and livestock ownership) [a-d] Capacity (access to information) [a-d] Agency (leadership roles) [a-d] Achievements (income, nutrition, food security, health, well-being) [a-d] Rating of technologies by group [a-d]	Variability and distributions resources, agency, and achievements [a-d]	[a] Key informant interviews [b] Participatory evaluation [c] Focus group discussions [d] Household survey
Social cohesion	Participation in community activities [a,b,c] Level and reliability of social support [a,b,c] Family cohesion [a,b,c]	Social groups [c] Participation in social groups [a,b,c] Incidence of social support [a,b,c]	[a] Household survey [b] Focus group discussions [c] Key informant interviews
Collective Action	Participation in a collective action group [a]	Collective action groups [a,b] Capacity of groups [a,b] Incidence of conflicts related to collective action [a,b] Effectiveness of conflict resolution measures [a,b]	[a] Household survey [b] Key informant interviews

Social Domain

Equity

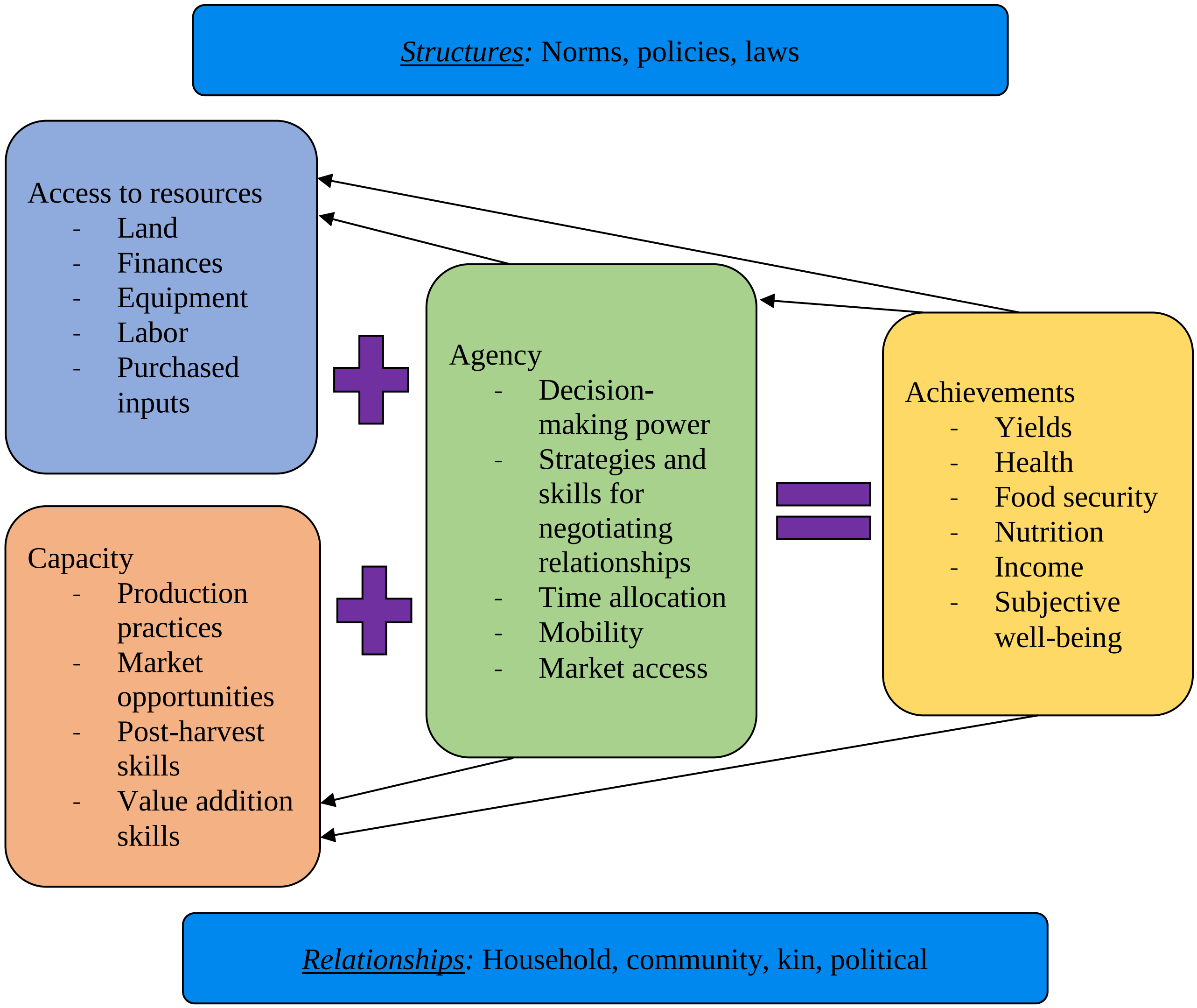
Description of the indicator

Equity is concerned with fairness or justice, which is a more complex concept than equality due to various ways that justice is understood. Cook and Hegtvedt (1983) outline four conceptions of justice: fair exchange, fair allocation, fair procedures, and just compensation. When equity is assessed across households, it can be easily calculated by comparing responses to household surveys; however, equity analyses for gender or youth require considering intra-household decision making as well.

Drawing from the gender empowerment literature, we developed a conceptual framework for equity in agriculture that is detailed in the figure below. Following Hemminger et al. (2014), we use the empowerment framework from Kabeer (1999) and included elements from the CARE gender toolkit and Women's Empowerment in Agriculture Index (WEAI) (Alkire et al., 2013) to categorize equity metrics as follows:

- Resources: These metrics are concerned with fair allocation of physical resources. They measure differential access to resources for agriculture.
- Capacity: These metrics are concerned with fair allocation of information and training resources. They measure differential access to information about markets or agricultural practices.
- Agency: These metrics are concerned with fair procedures. They measure differential levels of control over resources.
- Achievements: These metrics are concerned with fair exchange. They measure differences in realizing various benefits from agriculture.

Figure 1. Conceptual framework for empowerment and equality in agriculture. Adapted from Hemminger et al. (2014) and based on Kabeer (1999), including elements from CARE gender toolkit and Women's Empowerment in Agriculture Index (WEAI).



Quantitative assessment of equity can be calculated in various ways that combine the average values for each of the groups of interest (Group A and Group B). The following formulas are adapted from work by Rao (2016) on gender equity:

$$Parity\ =\ A/B$$

This is a ratio of the Group A to Group B measure, where a value of 1 signals no difference, <1 indicates B is favored and >1 indicates A is favored.

$$Gap\ =\ A-B$$

This is the difference between the Group A measure and the Group B measure, where a value of 0 signals no difference, <0 indicates B is favored, and >0 indicates A is favored.

$$Normalized\ Gap\ =\ 100*(A-B)/B$$

This is the percentage difference between the Group A and Group B measures, with the B measure as the denominator, where a value of 0 signals no difference, <0 signals Group B is favored, and >0 signals Group A is favored. The size of the differential is normalized against the Group B value.

When there are more than two groups of interest, then one group needs to be selected as the base for comparison (i.e., the Group B in the equations). The base for comparison may be the largest group, the average group, or the most favored group, depending on the situation.

Gender equity is a special type of equity that requires further attention due to the complexities of analyzing intra-household allocation, exchange, and procedures.

Most of the gender equity metrics require obtaining quantitative information from men and women and then calculating the gender gap. For easy interpretation, we follow the suggestion by Rao (2016) to compute the gender gap as the ratio of the female value to the male value expressed as a percentage, which we will refer to as gender parity:

$$Gender\ parity\ (\%)\ =\ female\ value/male\ value\ *\ 100$$

Ideally this calculation would be carried out using data collected separately from male and female adults in a given household. Interviewing multiple respondents per household is time consuming for data collection, entry, and analysis, but is likely to improve accuracy, especially where power struggles limit the free exchange of information or resources within the household. In some situations, it may be possible for one respondent to provide information about who in the household owns and controls various resources.

While this information is collected at the individual or household level it will often be useful to analyze it at the community level. In many cases, it may be useful to disaggregate the average gender indicators for various categories of women (e.g., household heads or parts of dual-headed household, junior or senior women in the household).

These equity equations provide a starting point for assessing quantitative data about equity that can be collected through interviews, focus group discussions (FGDs), or participatory exercises. However, equity assessment will be much richer and more informative if the quantitative values are accompanied by qualitative information about how and why the observed differences across groups came to be, and how fair or just the situation is perceived by various types of stakeholders.

For example, imagine two contexts where young adults age 17 to 30 have a significant gap in land ownership compared to middle-aged adults. In one context, the young adults may not feel it unfair to have less access to land because they are fairly compensated for their contributions while they wait to inherit. But in another context, young adults with the same gap in land access may have little hope of inheritance and feel that the status quo is unjust. Qualitative methods, such as FGDs and in-depth interviews, are critical for correctly interpreting these equity metrics.

Metric 1: Access to resources

Description of the metric

Land and livestock are critical resources for production and differences in ownership across groups and can reveal systemic inequities in how these resources have been allocated. Other key resources could be of interest in specific locations, such as irrigation water, credit, or machinery.

Measurement method 1: Surveys

Method of data collection and data needed to compute the metric

Household surveys are regularly used to collect information about land owned, land cultivated, livestock owned, and other agricultural resources (e.g, credit, machinery). Many baseline surveys collect this contextual data. Equity measures simply require the ability to disaggregate households and compare mean values (or distributions) among groups.

Open-ended questions can be part of a household survey where respondents are randomly selected; this enables the use of qualitative analysis to make inferences on the average perspective (i.e., most people from Group A feel that Group B is favored by the chiefs in land allocation) or to categorize perspectives by quantitative values, such as livelihood strategy, age, or wealth (i.e., most people who said that bribery was the main cause of unequal access to land were from poorer households and under the age of 30).

Example 1: Surveying access to land by gender

It is common to compare the average area of land used solely or jointly by women to the average area of land used solely or jointly by men. Due to the complexity of intra-household labor allocation, it is not possible to assume that those who work the land have the decision-making power about the benefits from their labor. We therefore suggest the use of the ability to decide how to use the harvest (sale or consumption) as a feasible metric for access to that land.

Where possible, land quality should be taken into consideration. For example, farmers' subjective assessment of soil fertility could be used to analyze the differences in quality of land that men and women have access to. The monetary value of the land would also show land quality, but accurately quantifying the market value for land is only possible where land markets are well developed.

Following Rao (2016) we focus on control over use of the harvest (e.g., home consumption, sale, trade). It is relatively simple in a household survey to add the question, "Who decides what to do with the harvest?" for each field, where multiple household members can be selected. Joint responsibility of a field should not be interpreted automatically as equity and will need to be interpreted in the local context.

Qualitative questions that could be useful for a deeper understanding of gendered responsibility include:

- In this community, on which fields do men do most of the work? On which fields do women do most of the work? On which fields do men decide what to plant and what inputs to use? On which fields do women decide what to plant and what inputs to use? Why?
- When someone says that they decide how to manage the harvest jointly as a household: What does that look like? How equal is the decision-making?

Example 2: Livestock ownership

Livestock ownership can either be separated by type of livestock (e.g., cattle, small ruminants, poultry) or combined using Tropical Livestock Units (Jahnke, 1982). In many agricultural surveys, the respondents are asked the number of all types of livestock. This could easily be followed up by the question, "Who is the owner of these livestock?" for each type. Asking about the monetary value for theoretical sale of each type of livestock could also allow for assigning quantitative values for comparing the value of livestock ownedamong groups.

Relevant questions adapted from Tanzania National Panel Survey (using the numbering system in that survey):

28. Who in your household decides what to do with these earnings?
29. In principle, who makes decisions about keeping or selling [ANIMAL]? (Indicate up to two people)
30. Who in your household provides labor for feeding/watering of [ANIMAL]?
31. Who in your household provides labor for selling the animals and animal products?
32. Who in your household mainly provides labor for grazing of [ANIMAL]?

Unit of analysis

Individual, household, or group

Limitations regarding estimating and interpreting

It should be noted that in some contexts, respondents may not truthfully reveal the quantity of land or livestock they own. For example, farmers with larger than average landholdings who are concerned about land redistribution may not mention all of the land that they own. On the other hand, respondents may exaggerate their livestock ownership due to its high social value.

Measurement method 2: Key informant interviews or focus group discussions

Method of data collection and data needed to compute the metric

In-depth interviews with key stakeholders from the various groups of interest or FGDs with members from these groups can provide approximations of resource allocation, as well as detailed information about how and why these resources are allocated. Open-ended questions are best for encouraging rich responses that draw on the respondents' lived experiences.

Qualitative interviews to understand equity in access to resources would aim to understand perceptions of the relative allocation across groups, how fair that allocation seems, and how and why there are differences. Purposively selecting respondents from various groups (or randomly selecting them from stratified lists) is important for qualitative methods regarding equity so that perspectives from all group of interest are obtained. In addition to selecting respondents from each group, in many contexts men and women will speak more freely in same-sex groups that are led by a facilitator who is of the same gender. FGDs may then be formed for Group A men, Group A women, Group B men, and Group B women. If age, wealth, or livelihood strategy is thought to be important then separate FGDs could be formed or the characteristics of the respondents in each different type of group, with the transcript designed to allow for analyzing differences in perspectives across these characteristics. Random sampling is not typically necessary with qualitative methods because statistical inference is rarely the goal.

Purposive sampling is important for targeting key informants with deep knowledge of a subject. Only well-informed individuals will be able to accurately estimate quantitatively the values of land and livestock resource owned by individuals in each group. For example, chiefs may have knowledge about how land has been allocated by ethnic group in their villages and the processes used for allocating that land.

During the interviews a secretary should take detailed notes and if possible the interview should be recorded so that respondents' exact words are the data that is analyzed. The transcripts from each interview or FGD should then be analyzed qualitatively. It is beyond the scope of this manual to detail the various forms of qualitative data analysis methods, but we will outline a basic strategy for categorizing information for a simple type of analysis:

1. Read through all of the transcripts and choose a few of the themes that you want to analyze in greater depth. These themes could come from your questions (e.g., response to how chiefs allocate land) or they could emerge from the responses to one or more of your questions (e.g., how migration to urban centers is differently affecting nomadic herders and settled farmers).
2. Copy all of the text relevant to one of your chosen themes into a single document. Include a respondent ID at the start of each portion of text so that you can easily identify who made each statement. If there is too much text to do this easily with copy and paste in a text editor then you can use qualitative software to code the data and then retrieve it by code. QDA Miner Lite is a free version of such software (https://provalisresearch.com/products/qualitative-data-analysis-software/freeware/).
3. Highlight the key words that relate to your theme in each response. You may consider using colored highlighting based on a group of responses or different colors for opposing responses (e.g., green for statements indicating the land allocation is fair and yellow for statements indicating it is unfair).
4. Summarize the diversity of responses in your own writing, – using quotation marks for direct quotes – aiming to fairly represent the breadth and depth of information as succinctly as possible. You may want to use numbers to represent the level of agreement on the statements (e.g., "…nine out of 12 respondents said…while the other three said…"). The written length of these summary statements depends on both the diversity of the responses being summarized and the detail necessary to achieve the purpose for which the summary is applied.

Unit of analysis

Individual or group

Limitations regarding estimating and interpreting

Qualitative interviewing and FGDs are time-consuming methods for collecting information about access to resources and typically are not used with hundreds of respondents. The purposive sampling strategy limits the usefulness of any quantitative data collected because inferences are limited to those similar to the respondents and leave the results open to critiques of selection bias. Analysis and/or quantification of the qualitative data can be overwhelming for scientists not trained in those methods, especially for assessing the complex causes and effects of unfair allocation of resources.

Measurement method 3: Participatory mapping and transect walks

Method of data collection and data needed to compute the metric

Participatory mapping and transect walks are activities that can be used with each group of interest (e.g., men, women, male youth, female youth) to better understand the resources that they use and have access to. For example, if the project focus is on crop production, then a map of the village farmland might be most appropriate. If the project focus is an irrigation scheme, then a transect walk through the irrigated land or along the canal might be useful.

Participatory mapping is typically done as a group on the ground, with local/available materials representing the features of the landscape. Once the main features are in place, probing questions can be used to add visual elements (e.g., placing different colored stones for land managed by men, women, and youth). The final map can be transferred to a large sheet of paper, which can then be easily copied for the community to keep and to be included in a report. Mapping can generally be employedto understand how resources are used by each group. Maps can also be used for specific planning or evaluation exercises, for example, deciding on a location for an investment in marketing, storage, or irrigation.

A variation on participatory mapping is the use of aerial or satellite images of the community – printed and laminated – for community members to draw on with different colored markers. This approach was used to apply a gender lens to nutrition sources in the landscape by Estrada-Carmona (2014). Another variation had men and women map where negotiations happened in decision-making (Christie and Luebering, 2011).

Transect walks are group walks across a landscape to observe the full range of conditions in an area (e.g., from low to high elevation). The walk does not need to be in a straight line but can meander to observe interesting elements. Someone should take notes about observations during the walk. At the end, the notes can be listed under a diagram or aerial map of the transect with images to represent the features along the route. For assessing access to resources, the walks should be done by separate groups and at various points the community members should point out what resources in the landscape they have access to and which ones they do not have access to. Further information on participatory mapping is available from Corbett (2009) and Willmer and Ketzis (2001).

Metric 2: Capacity

Method of data collection and data needed to compute the metric

Data is collected from a group of participants about the organizations, groups, and/or individuals that provide services to improve agriculture. One useful tool for the collection of this data is the gender-sensitive venn diagram from the Climate Change, Agriculture, and Food Security (CCAFS) gender toolkit (Jost et al., 2014).

Ask the participants to identify all of the organizations/groups/individuals, both local and external, that provide services and have an impact related to your area of focus (i.e., projects and activities that improve agriculture). Follow up by asking for a list of organizations/groups/individuals that are non-agricultural (e.g., finance, healthcare, women’s empowerment).

For each contributing organization, group, or individual that is mentioned, ask:

- What are the objectives of the organization?
- How long has the organization existed in the village?
- What has been its most successful project in the village? Why? Who benefited?
- Does this organization it have links with other/outside organizations? For what purpose?
- Who are the main beneficiaries? Men, women, young, old?
- Does one group (social and/or gender) rely more on the organization than others?

Transfer all of the organizations, groups, and individuals onto circles. The participants should decide which list items deserves small, medium, or large circles, to represent each organization’s relative importance to improved agriculture. Different colored circles can be used to indicate perceptions regarding groups or organizations. For example, green can be used to indicate organizations that the participants perceive to be friendly and easy to work with, while red is used for those that use too much scientific language or that do not seem to respect farmers.

Ask the participants if the organizations work together or have overlapping memberships. Leave the circles disconnected if they do not cooperate, use arrows if they only communicate, have them touch if they cooperate, have them overlap if they cooperate extensively.

Discuss the diagram with the following questions:

- Who holds decision-making roles in the organization?
- Does the organization have both men and women participating? If so, in what ways are they participating?
- Do women provide input in this organization? If so, how do the men react to it?
- Does the organization work specifically with women in agriculture or natural resource management?
- Does the organization provide information on farming practices? If yes, what is the nature of this information?
- Who accesses the information provided by the organization? Men? Women? How do they access it?
- Are the specific needs of young and elderly people taken into account by the institution? If so, how?
- Are the specific needs of marginalized groups, for example, ethnically, financially, socially marginalized, considered by the institution? If so, how?

For a variation on the Capacity activity, the UBALE gender analysis report (CRS, 2015) shows the use of a similar tool to explore women's access to services. Each service provider is listed and the size of the circle reflects the importance of the provider. The circles are then placed on a paper based on how accessible the services are to women, with the most accessible at the top and the least accessible at the bottom.

Metric 3: Agency

Description of the metric

Agency is the power that allows people to use the resources that are available to them to generate their desired achievements through decision-making, negotiation, deception, and/or manipulation (Kabeer, 1999). This is the element of empowerment that is most difficult to empirically observe. Leadership roles and formal decision-making authority can provide some indication of differential agency among groups.

Measurement method 1: Surveys

Method of data collection and data needed to compute the metric

Surveys can be used to identify the composition of leaders in groups, either by targeting key informants or by asking randomly selected individuals about the leadership of groups they participate in or are familiar with. Equal representation of groups in leadership positions does not necessarily mean there is equal agency.

Unit of analysis

Groups

Limitations regarding estimating and interpreting

Crafting a survey that effectively draws out respondents' experiences on a complex and relatively intangible concept can be difficult. Equal representation of groups in leadership positions does not necessarily mean there is equal agency.

Measurement method 2: Key informant interviews or focus group discussions

Method of data collection and data needed to compute the metric

As described in the Access to Resources Metric, qualitative interviews are suitable for understanding "how" and "why" things happen through the use of open-ended questions. In this case, qualitative methods can be used with respondents from various groups to obtain detailed information on the respondents' personal experiences in making decisions or being affected by the decisions of others, as well as their general perceptions about the decision-making power of various groups of interest. Crafting an interview guide that effectively draws out respondents' experiences on a complex and relatively intangible concept can be difficult.

Unit of analysis

Groups

Limitations regarding estimating and interpreting

Crafting an interview guide that effectively draws out respondents' experiences on a complex and relatively intangible concept can be difficult. Equal representation of groups in leadership positions does not necessarily mean there is equal agency.

Metric 4: Achievements

Description of the metric

The results of inequitable allocation of resources and/or disempowerment to use resources are observed as differential outcomes or achievements, such as different levels of income, nutrition, food security, and/or health or well-being. The methods for measuring these achievements are outlined in the indicators for other domains. The assessment of equity is primarily carried out by disaggregating the data into the groups of interest and then applying one of the three equations defined in the introduction to the equity indicator above.

Unit of analysis

Groups

Limitations regarding estimating and interpreting

Achievements often take time to materialize, even after underlying inequities or issues have been addressed. For this reason, care is necessary in interpreting the results. For example, equitable achievements may not reflect a recent trend of disempowerment. Likewise inequitable achievements do not mean that a short term project has failed, as those more equitable achievements may only gradually materialize.

Gender Equity

Description of the indicator

There are many tools available for carrying out gender analysis related to sustainable intensification of agriculture, many of which can be adapted for analysis of youth equity issues as well. Before reviewing those tools it is important to have a clear understanding of gender analysis. The following provides a clear summary:

> *Gender is the socially constructed difference between women and men. The meaning society gives to the roles of men and women results in certain power relations and dynamics. As a consequence, inequality in people's ability to make choices exists. Because women are often lagging behind in this respect, many tools are focused on empowering women. However, in order to change gender relations in society, the input of both men and women is required.* (Senders et al. 2012, p. 5)

It is also important to define a few key terms that will be used to discuss three distinct rights related to a resource (i.e., land): ownership, access, and control. "Ownership" of a resource refers to either having a legal title or having the right to transfer that resource to others. "Access" refers to the ability to use a resource, or "The freedom or permission to use a resource, perhaps with some decision making once access is obtained" (Pulhalla et al. [no date[, adapted from Feldstein et al. 1989). "Control" over a resource refers to the right to use and determine how the resource is used, or "The power to decide whether and how a resource is used, how it is to be allocated" (ibid).

Gender equity is a special type of equity that requires further attention due to the complexities of analyzing intra-household allocation, exchange and procedures.

Most of the gender equity metrics require obtaining quantitative information from men and women and then calculating the gender gap. For easy interpretation, we follow the suggestion by Rao et al. (2016) to compute the gender gap as the ratio of the female value to the male value expressed as a percentage, which we will refer to as gender parity.

$$\text{Gender parity (\%)} = \text{female value} / \text{male value} * 100$$

> This is a ratio of the female value to the male value times 100, where a value of 100 signals no difference, <100 indicates males are favored and >100 indicates females are favored.

This calculation ideally would be carried out using data from male and female adults in a given household. However, interviewing multiple respondents per household is time consuming for data collection, entry, and analysis. In some situations, it may be possible for one respondent to provide information about who in the household owns and controls various resources. While this information is collected at the individual or household level, it will often be useful to analyze it at the community level. In many cases, it may be useful to disaggregate the average gender indicators for various categories of women (e.g., household heads or part of a dual-headed household, junior or senior women in the household).

Metric 1: Land access by gender

Description of the metric

This metric compares the average area of land used solely or jointly by women compared to the average area of land used solely or jointly by men. Due to the complexity of intra-household labor allocation it is not possible to assume that those who work the land have decision-making power about the benefits from their labor; therefore, we suggest using the ability to decide how to use the harvest (sale or consumption) as a feasible metric for access to that land.

Measurement method: Mixed methods survey

Method of data collection and data needed to compute the metric

Following Rao et al. (2016) we focus on control over the use of the harvest (e.g., home consumption, sale, trade). It is relatively simple in a household survey to add the question "Who is responsible for deciding what to do with the harvest?" for each field, where multiple household members can be selected. Joint responsibility of a field should not be interpreted automatically as equity and will need to be interpreted within the local context. Qualitative questions that could be useful for a deeper understanding of gendered responsibility include:

- In this community, which fields are typically the responsibility of the man? Which fields are the responsibility of the woman? Why?
- When someone says that they decide how to manage the harvest jointly as a household, what does that look like? How equal is the decision-making?

Where possible, land quality should be taken into consideration. For example, farmers' subjective assessment of soil fertility could be used to analyze the differences in quality of land that men and women have access to. The monetary value of the land would also show land quality but accurately quantifying the market value for land is only possible where land markets are well developed.

Metric 2: Livestock ownership by gender

Description of the metric

Livestock ownership can either be separated by type of livestock (e.g, cattle, small ruminants, poultry) or combined using Tropical Livestock Units (Jahnke, 1982). In many agricultural surveys, the respondents are asked the number of all types of livestock. This could easily be followed up by a question, "Who is the owner of these livestock?" for each type of stock. Asking about the monetary value for each type of livestock if it were sold could also allow for combining livestock across categories.

Measurement method: Surveys

Method of data collection and data needed to compute the metric

Relevant questions from Tanzania National Panel Survey (World Bank, 2017):

28. Who in your household decided what to do with these earnings?
29. In principal, who is responsible for keeping [ANIMAL]? (Indicate up to 2 people)
30. Who in your household provided labor for feeding/watering of [ANIMAL]?
31. Who in your household provided labor for selling the animals and animal products?
32. Who in your household mainly provided labor for grazing of [ANIMAL]?

Metric 3: Time allocation by gender

Description of the metric

This metric can be used to assess gender equity through the quantitative measurement of differences in time spent on various tasks. While the division of labor by gender is not inherently negative, it is possible to assess gender equity by comparing amounts of leisure time for each gender or comparing time spent on the least desirable or most taxing tasks. Also, this information can be combined with other metrics in the agency and resource categories to assess who benefits from how the time is spent. Rao (2016) recommend the following metrics for gender labor inequities: “Average hours of leisure for women and for men or proportions of women and men who report inadequate leisure time”

Depending on the technology being assessed, it may be useful to develop detailed time allocation for activities directly or indirectly affected by that technology. In general, one can partition labor analyses into three broad categories—agricultural tasks (including livestock care), non-agricultural income generating tasks, household chores and leisure time. When inquiring about the time required for non-seasonal tasks, like household chores, it is common to ask about an “average” day. However, when inquiring about the time required for season tasks, such as crop production, there is no “average” day. Instead one can ask about all the agricultural activities (land preparation, planting, weeding, fertilizing, harvesting, etc.) field by field.

Measurement method 1: Daily time use exercise

CARE’s Daily Time Use exercise can be used to explore the differences between men’s and women’s daily activities, as described in the box below.

Box 1. CARE’s Daily Time Use exercise

- **Objective**: To explore and increase awareness of gender differences between women’s and men’s daily activities.
- **Materials/Preparation**: large sheets of paper, pens.
- **Participants**: This exercise has been used for both analysis and training.
    - Mixed groups of men and women in different ethnic/caste or socioeconomic groupings.
    - Single-sex groups.
    - Staff or research teams to critically reflect on gender roles.
    - Boys and girls, divided into single-sex groups.

Steps:

Following introductions and description of objectives, participants split into two groups by gender. The men’s and women’s groups separately list all the activities in their daily schedule, from waking to going to sleep. Each activity is drawn on an idea card and laid out in order across the day.

For this tool, it is important to specify what type of day is at the focus of the exercise, perhaps a day during the busiest time of the season, or after the harvest. The Exploring Dimensions of Masculinities exercise focuses on a typical weekday and a typical weekend day for its workshop with urban adolescent boys.

The group then reviews the day, and the facilitator discusses:

- Where does each activity take place? And with whom?

The facilitator then asks the group to identify which activity takes the most time. Next to that activity, the group places 10 stones. The group then identifies the second- most time-consuming activity, and decides how many stones to place there. This continues until each activity has stones next to it to show the amount of time required.

After this is completed, the facilitator asks the participants to list the activities across the daily schedule of someone of the opposite gender.

Once completed, the groups join together and the men and women each present their schedule to the whole group. When both schedules have been presented, the facilitator discusses:

- What surprised you about this exercise?
- Did the men accurately list women's activities? Did the women accurately list men's activities?
- Is there a difference in the kind of activities that men and women do? What is the difference?
- **Probe:** What is the reason for the difference? Does society expect very different things from men and women? Why does society expect men and women to spend time in different ways?
- **Probe:** Do you think this difference is justified? Why or why not?
- Which kind of work is a person paid for? Which kind of work is a person not paid for? Why?
- Which group has more leisure time to spend as they like? Which group has a larger workload?
- **Probe:** Is this justified? Why or why not?
- Was sex listed on the daily schedule? Why or why not? If it was added, would it be listed the same way in all the groups' daily activity schedules? Do men and women have the same expectations for sex? Why or why not?
- How much variation from this general daily activity schedule happens in your community? Do you see some particular men or women acting differently? Why is that?
- How does their reputation in the community change if they are not conforming to the norm?
- Are there certain ways that you would like to change community expectations of the daily activity schedules and workloads of men and women? What are they? Describe them. What can you do to make these changes happen? What can others do? How can this project contribute to those changes?

In addition, the facilitator may ask groups to place values alongside each activity:

- "H" (or another symbol) for tasks that are highly valued.
- "P" (or another symbol) for tasks that are paid with money.
- "R" (or another symbol) for those paid with respect/prestige.
- "U" (or another symbol) if it is unpaid.

Once completed, groups discuss their observations regarding the chart. The team then reflects on how the chart may change based on age or class. Teams may also discuss the roles of boys or girls in each of these tasks.

Variations:

Representing activities with tools: Rather than ask participants to draw, the facilitator can ask participants to gather the tools/utensils that they associate with each time period and lay them across a paper with the hours of the day to illustrate the activities. This is done with men and women side by side. The research team can then facilitate a discussion around the matrix and tools on the different daily activities done by men and women.

Activity pie chart: While the Daily Time Use exercise has been used with children, another variation asks separate groups of boys and girls to list the activities they undertake during the day. In the Power to Lead

Alliance, this exercise was facilitated with 10-11-year-old girls in one group and 12-14-year-old girls in another group.

This exercise begins with discussing the various activities or tasks that girls (or boys, in a separate discussion) do during the day. Make a list of the key activities together. Then invite one participant to draw a large circle on the ground or on a chalkboard. This circle represents a 24-hour day.

Explain that the group will now divide the circle into pie pieces, each representing one activity or task they have listed. The size of that piece should represent the time spent on that task. One way to represent the chart is to show or discuss what an orange looks like when it is cut into parts, with the wedges visible.

It may be helpful to start with the process by discussing how many hours of sleep girls get each night, and allocating that piece first.

Let the participants discuss and mark sizes themselves, as early as possible. The facilitator should focus on posing clarifying questions or probing for further discussion **(**e.g., "I see that this piece looks bigger than that one, so you spend more time fetching water than preparing dinner? Is this the same for everyone?"**)**

If the typical day is a school day, discuss how they spend their time in school (e.g., lessons, chores, meals, recess)

When the group finishes the chart, participants should review their list to be sure that each task has been included. They should note the amount of time allocated for each activity. Reviewing the list and adding the amount of time helps to confirm and clarify the pie drawing for analysis.

Following the activity, further discussion questions may include:

- If you were free to change your schedule, how would you spend your time differently?
- How might you work with others to change how you spend your time?

If separately performed with boys and girls, the groups can then be brought together in a subsequent activity to discuss and compare the two time-use pie charts.

Similarly, the CCAFS gender toolkit has a daily activity clock exercise. That manual suggests drawing two circles – one from 6 a.m. to 6 p.m. and the other from 6 p.m. to 6 a.m. – and then using the circles to map out pie-shaped pieces to represent time spent on various activities.

As Jost et al. (2014, p. 135) explains: "Be creative in how you use the different spaces (pieces of pie) to visually represent information. For example, once the clock is complete give the participants a pile of 100 beans to show the activities they feel are relatively more demanding or labor intensive compared to others. Or they can show the activities they find relatively enjoyable and rewarding compared to others. Or you can ask them to place a stone or other marker to show activities during which they obtain other benefits, like sharing information with others."

Measurement method 2: Time allocation 24-hour recall using individual surveys

The Women's Empowerment in Agriculture Index (WEAI) survey provides guidelines for estimating the number of hours worked per day in order to calculate its time allocation component of the empowerment index (Alkire et al., 2013). It recommends asking people about how they spent their time from 4:00 a.m. the previous day to 3;59 a.m. the next morning. The various activities are predefined and can be recorded in 15-minute intervals. Respondents can provide up to two activities at any one time, but will be asked which activity is primary. Following is the first page of the table for data collection. The second page has the same activities for the remaining hours of the 24-hour period.

Table 3. Time allocation data sheet from the WEAI

MODULE G6: TIME ALLOCATION

Enumerator: ***G6.01:*** *Please record a log of the activities for the individual in the last complete 24 hours (starting yesterday morning at 4 am, finishing 3:59 am of the current day). The time intervals are marked in 15 min intervals and one to two activities can be marked for each time period by drawing a line through that activity. If two activities are marked, they should be distinguished with a P for the primary activity and S for the secondary activity written next to the lines. Please administer using the protocol in the enumeration manual.*

		Night		Morning				Day					
	Activity	4	5	6	7	8	9	10	11	12	13	14	15
A	Sleeping and resting												
B	Eating and drinking												
C	Personal care												
D	School (also homework)												
E	Work as employed												
F	Own business work												
G	Farming/livestock/fishing												
J	Shopping/getting service (incl health services)												
K	Weaving, sewing, textile care												
L	Cooking												
M	Domestic work (incl fetching wood and water)												
N	Care for children/adults/elderly												
P	Travelling and communiting												
Q	Watching TV/listening to radio/reading												
T	Exercising												
U	Social activities and hobbies												
W	Religious activities												
X	Other, specify…												

Limitations regarding estimation and interpretation

Alkire et al. (2013) note that a major shortfall of the WEAI method is that it does not cover seasonality and may not be representative of the given season if the previous day was a holiday. Harvey and Taylor (2000) recommend asking respondents about activities that took place no more than one or two days previous to the interview, as memory fades on the detailed use of time beyond that.

Harvey and Taylor (2000) also point out the importance of adapting all these tools to the local context using local terms for how time is reckoned. In addition, they note that data generated from questions asking only about specific activities (e.g., carrying water, weeding, collecting firewood) are susceptible to significant reporting errors compared to more rigorous methods like the stylized activity log used in the WEAI time module.

Measurement method 3: Activity analysis

The goal of this exercise is to understand who does each activity. This can be a binary question for each gender (i.e., yes or no) or it could be a proportional assessment of time spent on each task (i.e., percent, or proportionally allocating 10 stones by gender). This analysis could be implemented through individual interviews, couple interviews, or larger groups (e.g., several couples, separate groups of women and men).

Table 4. Activity analysis data table

	Men	Women	Boys	Girls	Comments

Crop/Field 1					
Activity 1:					
Activity 2:					
Activity 3:					
Crop/Field 2					
Activity 1:					
Activity 2:					
Activity 3:					
Livestock – Animal 1					
Task 1:					
Task 2:					
Task 3:					
Household production					
Off-farm production					

Measurement method 4: Gender-sensitive seasonal calendar

In order to collect this information, symbols can be used for each gender, tasks from the activity analysis above can be referenced, or a separate calendar can be done for men and women.

Table 5. Seasonal activity data sheet

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Season												
Crops												
Livestock												
Household												
Off-farm												

Metric 4: Management control by gender

Description of the metric

This metric aims to capture differences in decision-making power between men and women. To be operationalized, it will be necessary to choose the most important decision within the given context. For cropping systems, one could measure the land area where women report that they are the primary decision maker about crop management (solely as well as jointly) compared to the land area where men report being the primary decision maker (solely as well as jointly). Some simple survey questions are “Who decides what crop to plant?” “Who decides what inputs to apply?” and “Who decides when to plant, fertilize, or weed?” Agency over the use of production factors (i.e., plowing) can be indirectly measured by asking, for example, when women’s and men’s fields get prepared. Management control gender gaps need to be explored for other areas of agriculture as well, such as livestock raising, irrigation schemes, and collective marketing efforts.

A longer-term focus for cropping systems may be helpful where SI technologies aim to improve land quality. Ownership of land shows that the individual has the incentive to make long-term investments; however, quantifying ownership of land by gender is not a simple matter for two reasons. First, ownership of land may mean different things in different contexts. Complete ownership would include having the right to manage it, the right to control the benefits from it, and the right to transfer rights to others (Rao, 2016). In many developing country contexts, traditional tenure systems do not give individuals the rights to transfer land, so the term “ownership” refers simply to the rights to manage and to control benefits from the land. Second, quantifying ownership is difficult because de facto ownership may be different from de jure ownership (i.e., the name on the title). Rao (2016) justifies a focus on de facto rights by giving the example of someone officially owning distant land that they are not able to access, while another has access to land without a title. These de facto rights to land need to be assessed at the individual level and not simply at household level.

Measurement method 1: Intra-household decision-making tool

CARE's Intra-household decision-making tool can be used to understand how decisions are made between men and women in a household, as described in the box below.

Box 2. CARE's Intra-household decision-making tool

- **Objective**: To understand how decisions are made around resources and strategies women use to influence men's decision-making.
- **Materials/Preparation**: Interview checklist based on relevant literature and discussions with field staff and partners familiar with the local context. Teams should also discuss the translations for "power" and "empowerment" to be used with respondents. If time permits, interviews should be piloted and adjusted before the study itself, and adapted for interviews with men. To prepare staff for research, teams in CARE Bangladesh also conducted mock interviews.
- **Participants**: Men and women across age, household composition, ethnicity, and well-being groups in individual interviews.

Steps:

This research uses semi-structured interviews and key personal events in women's lives (i.e., education, dowry, marriage, work/income for women and their daughters) as the basis for discussing decision-making, women's interpretation of events, and women's use of power.

Sample questions from the Tanzania National Panel Survey:

Education:

- How much education of self/spouse? Literacy?
    - *Probe:* Why did(n't) you go to school?
- Education of children? Education plans for boy children? And for girl children?
    - *Probe:* Barriers? Factors preventing education?
- In your opinion, why is education important?
- Without education, where do you see the future of your children? Why?
- With education, where do you see the future of your children? Why?
- In your family, how is the decision made whether children go to school or not?

Marriage:

- When were you married and tell me about the circumstances (Your age and that of your spouse; who made the decision; bride-price).
- What are you planning/what happened for your children? (same kind of ideas as above).
    - *Probe:* (if different) Why was this different?
- Do you practice any form of birth control? What kind? Is this a collective decision with your husband/partner? Is it a personal decision?
- [*if 1st wife in polygamous marriage*]: Were you consulted in the second marriage? How has your life changed since the arrival of a new wife?
- [*if polygamous man*]: Did you consult your 1st wife in your second marriage?
- [*if inherited/widow*]: Please share with us the circumstances after your husband's death.

- [*if divorced/living with partner/abandoned/FHH*]: Please share with us the circumstances of your ________. Probe about: social support, stigmatization, etc.

Gender roles in the household:

- What is your responsibility inside the household (i.e., cooking, childcare, domestic duties, crop processing)? Was it the same for your mother or different?
- What are the responsibilities of your spouse or partner? Was it the same for your father or different?

Decision-making:

- Mobility
  - How far away is your original family?
  - How often do you visit your birth village? And how often are you free to go?
  - [for *women*] What is the farthest you have been away from your home? And with whom?
  - Are you free to go anywhere or do you have to consult first? What are the circumstances (doctor, visiting family, visiting friends, markets)?
- Assets and income
  - Are you a member of a Village Savings and Loan group? If yes, how much do you contribute on a weekly basis? Have you taken any loans and how did you spend the money?
  - [for *women*] Do you earn any cash income?
  - [for *women*] Do you own any assets? (livestock, goats, ducks, chickens)
  - [for *women*] Do you own land? Do you rent in land? Do you cultivate any land where the crop is yours? Is there any petty trading? (i.e., burning charcoal)
  - [for *women*] Do you ever have money of your own where you can decide how to spend it?

Conclusion:

- When have you felt most in control of your life? When have you felt least in control of your life?
- Who is your role model? And why?

As preparation, a mock interview between a researcher and a field facilitator can be observed by the other facilitators, who then can carry out the interviews. For best responses to the interviews, women interview women and men interview men.

Metric 5: Market participation by gender

Description of the metric

Within a household, this could be a comparison of who markets which products. At the landscape scale, the incidence of men and women participating in the market could be compared.

Measurement method 1: Gender-focused value chain analysis using focus group discussions

Method of data collection and data needed to compute the metric

The "gender in value chains toolkit" developed by Agri-ProFocus includes instructions and examples of questions for FGDs and mapping exercises to analyze the role of women in particular value chains and how to "make visible" their contributions, even in value chains thought to be dominated by men.

Tool 3.2a, "Making a Gender-Sensitive Value Chain Map," has the following steps:

- Formulate hypotheses about women's roles and possible entry points for greater participation in the value chain.
- Actor mapping – make a visual presentation of the actors along the value chain by gender.
- Make invisible women stakeholders visible – think through the roles women play in each step, even if indirectly.
- Activity mapping – distinguish gendered roles for each activity along the value chain.
- Specific gender mapping – document how many actors and jobs (disaggregated by gender) are involved in each stream of the value chain (e.g., informal sector vs. industrial sector for milk production).

Metric 6: Income by gender

Description of the metric

Income is both a resource for and an achievement from women's empowerment. When considering it as a resource, the focus is on access to finances and can be measured by asking who participates in the decisions to buy items such as agricultural inputs and daily goods. When considering income as an achievement, it can be measured based on net income from crops or animals controlled by each gender. If detailed time allocation has been collected, then returns to labor can be calculated and compared across genders.

Metric 7: Nutrition, food security, and health by gender

Description of the metric

These metrics use disaggregated data from the Human Condition Domain to compare achievements across gender.

Metric 8: Ratings of technologies by gender

Description of the metric

Technologies that are used at the farm and field scale may be evaluated differently by men and women. The data collection happens at the household scale so the gendered rating is listed at the household level.

Measurement method 1: Participatory rating

Method of data collection and data needed to compute the metric

Preparatory information: Ask participants to identify the criteria affecting their decision whether or not to use an agricultural innovation/practice. For example, if it is a new bean variety, what characteristics do they look for in beans? Have them vote to prioritize that list of criteria by providing three to five votes for each person to distribute among the criteria.

Rating an innovation: Ask the participants to rate the practice/innovation according to the most commonly listed criteria. Some participants may be comfortable with numbers and giving an innovation a score, such as from 1 to 5 or from 1 to 10. Two methods to make it easier for participants to accomplish the ratings:

- Option 1: Putting a marker on a line. To use this method you should prepare a sheet with a line marked for each rating and a symbol on either end (for example a horizontal line with five vertical crosses representing 1 to 5, and if assessing harvest using the symbol of a small sack at level 1 and a pile of full sacks at level 5). A separate sheet could be used for each criterion, a blank sheet can be reused for each, or a laminated sheet can be drawn on and wiped off. To compare multiple innovations or practices, create a marker for each (such as a drawing on a small piece of sticky paper) and have them place each marker on the line.
- Option 2: Draw a matrix on the ground with the criteria across the top and the innovations/practices to be compared as the rows. Give the participants a pile of markers (e.g., stones, beads, seeds) and ask them to put up to five markers in each square to evaluate the importance of each innovation.

Measurement method 2: Co-benefit analysis

Method of data collection and data needed to compute the metric

This method is described in the CCAFS gender toolkit. It uses FGDs with separate groups of men and women in order to understand the perceived benefits and burdens from various agricultural practices.

Box 3. Co-benefit focus group analysis process

Steps:

1. Begin the discussion by asking about one agricultural or nonagricultural change that is of interest. Probe the focus group to understand the different benefits and burdens from each practice.
2. Once lists of benefits and constraints have been noted, ask a volunteer to list or draw them out on many sheets of paper or on a large poster.
3. Take 100 counters or beans and explain that they represent all of the men or women (depending on the disaggregated group). Ask a volunteer to distribute the counters by importance among the benefits from the agricultural practice. Encourage the group to work together to create a distribution upon which they agree.
4. Repeat this step, but distribute the counters by importance among the burdens of adopting the practice (i.e., the worst burdens will get the most counters).
5. Discuss the results as a group to gain more insight about the perceptions of the benefits and burdens of the change.
6. Follow the same process for each practice of interest. Record the benefits, burdens, and discussion notes for each practice.
7. Compare results from men and women.

Probing questions:

- How does this activity affect soil quality?

- How does this activity affect water sources?
- How does this activity affect forest resources?
- How does this activity affect crop diversity?
- How does this activity relate to land tenure? Is land required? Rented? Shared in common? Privately owned?
- Who has control over land? Who has access to land? How does those who do not own land gain access to it?
- How is the burden of labor for this activity shared? Who does most of the work? Is it done in a group?
- Does this activity require buying or renting of equipment? Can all groups or individuals in the village afford the equipment? If not, how is it shared? Who cannot afford it?
- Are there seasonal or time constraints associated with the equipment? Who operates the equipment? Who rents it?
- How time consuming is this activity? How does it affect amount of labor for men? For women? For children?
- Is there special knowledge required to do this activity? Who holds this knowledge? Who does not?
- How does this activity effect household food security or consumption?
- Does this activity have any nutritional benefits? Who makes the decision to invest in nutrition? Who in the family does it benefit the most in terms of nutrition?
- How does this activity affect overall family income? Who keeps the income? Is it shared?
- Is the income from this activity channeled into long-term investments like education, businesses, or loan repayment? Who makes the decision to invest? Who benefits most?
- How is information shared within a group or household engaged in this activity or among individuals?
- Are there small businesses that have grown from this activity? Do men, women, or children run these businesses? Are there associations that run the business? Is the membership of associations mostly men or women? How are decisions made in associations? How are benefits shared?

Metric 9: Women's Empowerment in Agriculture Index (WEAI)

Description of the metric

The WEAI (Alkire et al., 2013) is calculated by following a specific data collection method where male and female responses are compared. This survey process may be too demanding for many programs, but it does provide a great deal of information about the various facets of empowerment at the community or regional scale.

Method of data collection and data needed to compute the metric

The WEAI has five domains for the empowerment subindex: Production, Resources, Income, Leadership, and Time.

Table 6. Domains of the WEAI Empowerment subindex

Domain	Indicators	Weight	Abbreviated weight
Production	Input in productive decisions	1/10	1/5
	Autonomy in production	1/10	
Resources	Ownership of assets	1/15	2/15
	Purchase, sale, or transfer of assets	1/15	
	Access to and decisions on credit	1/15	1/15
Income	Control over use of income	1/5	1/5
Leadership	Group membership	1/10	1/5
	Speaking in public	1/10	
Time	Workload	1/10	1/5
	Leisure	1/10	

The gender parity subindex of the WEAI is calculated from two components:

- Gender parity – the percentage of women who have gender parity defined as either being empowered (scoring 0.80 or higher on the empowerment subindex) or having an empowerment index score greater than that of the primary male in their household.
- Empowerment gap – the average percentage shortfall that a woman without parity experiences relative to the male in her household.

The questions in the table below are the core questions from the WEAI survey. They provide a concise summary of the information that can be gained from implementing the WEAI; however, when actually carrying out the survey it is important to change the order of the questions and to use additional questions to improve the flow. A formatted questionnaire should be developed to help the enumerators ask the questions and mark the responses, and response codes should be designated to facilitate data entry and analysis.

Table 7. Core WEAI survey questions by dimension

Production	1. How much input did you have in making decisions about: food crop farming, cash crop farming, livestock raising, fish culture?
	2. To what extent do you feel you can make your own personal decisions regarding these aspects of household life if you want(ed) to: agriculture production, what inputs to buy, what types of crops to grow for agricultural production, when or who would take crops to market, livestock raising?
	3. My actions in [DOMAIN: agricultural production, inputs to buy, crops to grow, take to market, livestock] are partly because I will get in trouble with someone if I act differently.
	4. Regarding [DOMAIN] I do what I do so others don't think poorly of me.
	5. Regarding [DOMAIN] I do what I do because I personally think it is the right thing to do.
Resources	6. Who would you say owns most of the [ITEM]? Agricultural land, large livestock, small livestock, chicks, etc.; fish pond/equip; farm equip (non-mech); arm equip (mechanized); nonfarm business equipment; house; large durables; small durables; cell phone; non-ag land (any); transport.
	7. Who would you say can decide whether to sell, give away, rent/mortgage [ITEM] most of the time?
	8. Who contributes most to decisions regarding a new purchase of [ITEM]?
	9. Who made the decision to borrow/what to do with money/item borrowed from [SOURCE]? nongovernmental organization (NGO); informal lender; formal lender (bank); friends or relatives; ROSCA (savings/credit group)
Income	10. How much input did you have in decisions on the use of income generated from: food crop, cash crop, livestock, non-farm activities, wage & salary, fish culture
	11. To what extent do you feel you can make your own personal decisions regarding these aspects of household life if you want(ed) to: Your own wage or salary employment? Minor household expenditures?
Leadership	12. Are you a member of any: agricultural/livestock/fisheries producer/mkt group; water; forest users'; credit or microfinance group; mutual help or insurance group (including burial societies); trade and business association; civic/charitable group; local government; religious group; other women's group; other group
	13. Do you feel comfortable speaking up in public: To help decide on infrastructure (like sm wells, roads) to be built? To ensure proper payment of wages for public work or other similar programs? To protest the

	misbehavior of authorities or elected officials? To intervene in case of a family dispute?
Time	14. Workload based on activities—see 24-hour recall question in the time allocation metric
	15. How would you rate your satisfaction with your available time for leisure activities like visiting neighbors, watching TV, listening to radio, seeing movies, or doing sports?

Below is an example of how WEAI questions 1 and 10 from the above table are implemented in a survey, with all of the features mentioned, such as question and response guides and spaces to mark coded responses.

Table 8. Example WEAI question instrumentation

MODULE G2: ROLE IN HOUSEHOLD DECISION-MAKING AROUND PRODUCTION AND INCOME GENERATION

Household identification (in data file, each sub-module (G2-G6) must be linked with HH and respondent ID)

Respondent ID Code

Activity		Did you (singular) participate in [ACTIVITY] in the past 12 months (that is during the last [one/two] cropping seasons)? Yes......... 1 No.......... 2 >> next activity	How much input did you have in making decisions about [ACTIVITY]?	How much input did you have in decisions on the use of income generated from [ACTIVITY]
ActivityCode	Activity Description	G2.01	G2.02	G2.03
A	Food crop farming: crops that are grown primarily for household food consumption			
B	Cash crop farming: crops that are grown primary for sale in the market			
C	Livestock raising			
D	Non-farm economic activities: Small business, self-employment, buy-and-sell			
E	Wage and salary employment: in-kind or monetary work both agriculture and other wage work			
F	Fishing or fishpond culture			
			G2.02/G2.03: Input into decision making No input.............................. 1 Input into very few decisions.............. 2 Input into some decisions.................... 3 Input into most decisions..................... 4 Input into all decisions......................... 5 No decision made.............................. 6	

Social Cohesion

Description of the indicator

Direct indicators of social cohesion are “membership rates of organizations and civic participation” and “levels of trust” (in other people), while proxies for social cohesion are income distribution and ethnic heterogeneity (Easterly et al., 2006). Social cohesion is seen as society level issue, while social capital is micro-level issue. Grootaert et al. (2004) have developed a guide for measuring social capital with household surveys.

Collective Action

Description of the indicator

Collective action is common in many areas for managing natural resources (e.g., irrigation, water, fisheries). Collective action can also refer to cooperative efforts in agriculture for marketing, processing, procuring inputs, etc. Collective action can be affected by changes in the community if that change alters incentives or affects trust levels.

References

Adolwine, W.M., and A. Dudima. 2010. Women's access to emerging urban land in the Sissala East District in Northern Ghana. Journal of Science and Technology 30(2): 94–104.

African Development Fund (ADF). 2005. Republic of Malawi multi-sector country gender profile.

Alkire, S., R. Meinzen-Dick, A. Peterman, A. Quisumbing, G. Seymour, and A. Vaz. 2013. The women's empowerment in agriculture index. World Development 52: 71–91.

Apusigah, A.A. 2009. The gendered politics of farm household production and the shaping of women's livelihoods in Northern Ghana. Feminist Africa (12): 51–68.

Britwum, A.O., N.K. Ghartey, and P. Agbesinyale. 2006. Organising labour in the informal sector: the case of rural agriculture in Ghana. Ghana Universities Press, Accra, Ghana.

Britwum, A.O., D. Tsikata, A.D. Akorsu, and M.A. Aberese. 2014. Gender and land tenure in Ghana: a synthesis of existing literature. ISSER Technical Publications No 92, Legon, Ghana.

Britwum, A.O. and A.D. Akorsu. 2016. Qualitative gender evaluation of agricultural intensification practices in northern Ghana. Ibadan, Nigeria, International Institute of Tropical Agriculture. Africa RISING consultancy report. http://hdl.handle.net/10568/78479

Chirwa, W.C. Mvula.; P.M.; Dorward, A. and M. Matita. (2011). "Gender and Intra-Household Use of Fertilisers in the Malawi Farm Input Subsidy Programme". Future Agricultures Working Paper No: 028.

Chitsike, C., J. Kampanje-Phiri, and N. Sangole. 2016 (December draft). A qualitative evaluation of gender aspects of agricultural intensification practices in Central Malawi. Africa RISING consultancy report. International Institute of Tropical Agriculture, Ibadan, Nigeria.

Catholic Relief Services (CRS). 2015. UBALE (United in Building and Advancing Life Expectations). Gender Analysis Report. CRS, Lilongwe, Malawi

Christie, M.E. and C. Luebering. 2011. Women as mapmakers: gender and empowerment in participatory mapping. Paper presented at the Annual Conference of the Association of American Geographers on 14 April 2011 in Seattle, WA, USA. Retrieved from: https://vtechworks.lib.vt.edu/handle/10919/70232

Cook, K.S. and K.A. Hegtvedt. 1983. Distributive justice, equity, and equality. Annual Review of Sociology 9(1): 217–241.

Corbett, J. 2009. Good practices in participatory mapping: A report prepared for IFAD. Retrieved from: http://www.ifad.org/pub/map/PM_web.pdf

Duncan, B.A. 2004. Women in agriculture in Ghana. Friedrich Ebert Foundation, Accra.

Duncan, B.A. and C. Brants. 2004. Access to and control over land from a gender perspective: a study conducted in the Volta Region of Ghana. FAO, Accra.

Egyir, I.S., C. O'Brien, and J. Bandanaa. 2017. Gender equality policy approach to food security in Ghana: findings from a USAID Feed the Future Project. World Medical & Health Policy. In Print, manuscript ID WMHP-12-16-007.

Estrada-Carmona, N. 2014. Using participatory mapping with a gender lens to understand how landscapes are used for nutrition. Accessed 11 May 2017 from: http://a4nh.cgiar.org/2014/11/06/participatory-mapping-with-a-gender-lens/

Feldstein, H., S. Poats, and K. Cloud.1989. Conceptual framework for gender analysis in farming systems research and extension. Pages 7–26 in Working Together Gender Analysis in Agriculture. Kumarian Press, West Hartford, Connecticut, USA

Food and Agriculture Organization of the United Nations (FAO). 2011. Gender inequalities in rural employment in Malawi: an overview. Prepared by the Gender, Equity and Rural Employment Division of FAO. United Nations, Rome.

Harvey, A.S. and M.E. Taylor. 2000. Time Use. Chapter 22 in Designing household survey questionnaires: lessons from fifteen years of the living standards measurement study, volume 2, edited by M. Grosh and P. Glewwe. The World Bank, Washington, D.C., USA.

Hemminger, K., B. Bock, J. Groot, M. Michalscheck, and C. Timler. 2014. Towards integrated assessment of gender relations in farming systems analysis. MSc thesis, Wageningen University (also presented as a poster with the co-authors at Tropentag 2014).

Hockett, M. and R. B. Richardson. (2016). "Examining the Drivers of Agricultural Experimentation Among Smallholder Farmers in Malawi". Expl Agric: pp1-21, Cambridge University Press.

Hyder, A. and J.R. Behrman. 2014. Female economic activity in rural Malawi. Journal for Development and Leadership 3(1), p.1-10.

Jahnke, H.E. 1982. Livestock production systems and livestock development in tropical Africa Vol. 35. Kiel: Kieler Wissenschaftsverlag Vauk.

Jost, C., N. Ferdous, and T.D. Spicer. 2014. Gender and inclusion toolbox: participatory research in climate change and agriculture. CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), CARE International and the World Agroforestry Centre (ICRAF). Copenhagen, Denmark. Available online at: www.ccafs.cgiar.org

Kabeer, N. 1999. The conditions and consequences of choice: reflections on the measurement of women's empowerment (108: 1–58). United Nations Research Institute for Social Development (UNRISD), Geneva.

Koira, A.K. 2014. Agribusiness in sub-Saharan Africa: pathways for developing innovative programs for youth and the rural poor. Mastercard Foundation Working Paper.

Manuh, T., J. Songsore, and F. Mackenzie. 1997. Gender and land: the interface between legislative initiatives, customary tenure and land use management, mimeo.

Mathiassen, A., R. Eliasi, M. Mahowe, D. Chunga, G. Iversen, S. Pedersen and D. Roll-Hansen (2007), Gender assessment for Malawi. Statistics Norway Report No.25. Oslo: Statistics Norway.

Ministry of Food and Agriculture. 2007. Food and Agriculture Sector Development Policy (FASDEP II). MoFA, Accra, Ghana.

Ohene, V. 2013. Determinant of farmers' participation in the Youth-in-Agriculture Programme in the Eastern Region of Ghana. MPhil. thesis submitted to the University of Ghana, Legon, Accra.

Pulhalla, J., V. Michener, and E. Bastidas. No date. Conceptual framework for gender analysis in farming systems research and extension. University of Florida Digital Collections: http://ufdc.ufl.edu//AA00008143/00001

Rao, S. 2016. Indicators of gendered control over agricultural resources: A guide for agricultural policy and research. Working Paper No. 1. CGIAR Gender and Agriculture Research Network, CGIAR Consortium Office and International Center for Tropical Agriculture (CIAT). Cali, Colombia.

Senders, A., A. Lentink, and M. Vanderschaeghe. 2012. Gender in value chains—practical toolkit to integrate a gender perspective in agricultural value chain development. Published by Agri-ProFocus, available from: http://www.fao.org/sustainable-food-value-chains/library/details/en/c/265588/

Simtowe, F.P. (2010). "Livelihoods diversification and Gender in Malawi". African Journal of Agricultural Research 5(3): 204 – 216.

United States Agency for International Development (USAID). 2010. Malawi Country Profile: Property Rights and Resource Governance. USAID, Lilongwe, Malawi 25pp.

Willmer, A. and J. Ketzis. 2001. Participatory gender resource mapping: a case study in a rural community in Honduras. PLA notes 33. IIED, p. 17-22.

World Bank 2017. Living Standard Measurement Study Survey Integrated Surveys on Agriculture (LSMS-ISA).