V. MEASUREMENT PROBLEMS IN FIELD WORK

A. Some Sources of Measurement Problems-Review

1. Respondent don’t know answer
   (i.e., continuous, non-registered data)

2. Respondent knows “answer” but only in imprecise terms
   (e.g., local measure hard to convert consistently across
   respondents, “cigarette tins”, “bags”)

3. Respondent’s recall is poor
   (e.g., continuous harvesting of “green” corn, cassava)

4. Respondent doesn’t want to reveal sensitive data
   (e.g., culture—“concept of the limited good”: farmer believes that if
   others learn of his/her success/high yield, it will bring bad luck)
Consider: Do you really need highly accurate data?

Decision to try to obtain highly precise data (via direct measurement) should be based on:

- Accuracy required to achieve your research objectives
- Key variables in your proposed be analyzed (e.g., may need to only obtain accurate data for a few variable
- Cost/time required to directly measure the phenomena
- Availability of resources (e.g., skilled staff, measurement tools)

B. Land Area Measurement

1. Possible reasons for wanting precise estimates of land area

   - Want to estimate crop yields with a high degree of precision, so you need an accurate land area estimate (i.e., yield=production divided by planted area)

   - Want to estimate wealth (i.e., in rural areas, land area is a good indicator of wealth)

   - Land area proxies are highly unreliable (e.g., seeding rate, farmers’ “guesstimates” are unreliable)
2. Direct measurement difficulties (sources of non-sampling error)
   • Irregular-shaped fields (e.g., slash & burn/shifting agriculture)
   • Farmer may plant several fields, so must measure each field
   • Fields may be subdivided into parcels with different technologies, so must measure each parcel
     (def. A parcel: single crop, variety, land preparation method)
   • If fields are remote, may require much time to visit each field
   • Area planted may be > area harvested, so must decide which you want to estimate/measure? If harvested area, must measure near harvest time
   • Area planted may vary by cropping season/year, but can’t measure for past seasons/years (e.g., may “rent out” in wet season, plant in dry season)

3. Direct ways to measure area
   a) Must define the holding (definition?) to be measured
      • Area planted or harvested?
      • Area owned or area used this year? (e.g., owned + borrowed - rented out)
      • Which seasons?
      • Farmer’s land or family’s total land?

   • Non-planted area may be large, so may need to to subtract out area in bunds, canals, fallow land in the field
   • Many crops may be grown in the same field, making it hard to estimate area planted to a specific crop (Mazuze)
b) Triangulation (Figure)
Def. Visit the field and estimate the area by: dividing the field into triangles, measure angles & distance between two points, apply trigonometry formulas—or use a GPS unit to measure

- Most frequently used method, but time consuming

c) Aerial Photographs
- Most frequently available for irrigation systems

- If available, may be difficult to delineate farmer’s fields/boundaries, especially for “parcel” subdivisions

d) If field is inter-cropped (e.g., corn + beans), may be impossible to directly measure each crop’s area. Can report area as a crop mix, and then estimate mix percentage
(e.g., 4 rows of maize, rows of beans=66% maize, 33% beans)

4. Indirect Alternatives for measuring land area (proxies)

- Amount of seed planted. But, must confirm the conversion factor (seed/ha) that farmers are reported to use
(e.g., directly measure several farmers’ fields and divide the area by the amount of seed the farmers planted to estimate/confirm the reported seed rate/ha)

- Number of mounds/hills planted
(e.g., for yams, cassava; but must also estimate hill spacing)

- Number of trees
(e.g., coffee, bananas; but must also estimate tree spacing)

5. But, farmers’ farm/crop area estimates are generally OK, especially if land is titled & crops are planted as a monocultures. However, you should assess the accuracy of farmers’ estimates by asking farmers how they estimate their farm/crop area

- Mazuze-sweet potato area measured as equivalents to HH lot
C. Crop Production (and Yield) Measurement

1. Possible reasons for wanting precise production/yield estimates

   • Need to accurately estimate farm income
     (i.e., production is the primary source of HH income in rural areas)

   • Yield is key dependent variable in a technology impact study
     (e.g., want to estimate yield impact of oxen vs. tractor plowing, traditional vs. improved varieties)

   • You plan to estimate a production function, which requires highly precise production (and crop area) data
     (e.g.,  \( Y = f(x_1 \ldots x_n) \))

2. Direct measurement difficulties (sources of non-sampling error)

   • Measuring yield includes all area measurement problems
     (i.e., yield = production/area)

   • Part of crop may be home-consumed/harvested early, so HH may not consider it part of the harvest
     (e.g., especially crops harvested over time, like “green” corn, cassava)

   • Some crop may be continually harvested, so it difficult for farmers to recall total production
     (e.g., cassava, climbing beans, cucumbers)

   • If crop is intercropped, must convert field yields to pure stand yield
     (e.g., FAO’s low yield estimates for LDCs are likely due to intercropping)
• Biological yields (estimated by a crop cut) are often not the same as farmers' actual economic yield
  o Harvesting losses may approach 10%
  o Part of the field may be in nonproductive area (bunds); so yields estimated from crop cuts are > farmers’ actual yield

3. Direct measurement methods
   a) Yield plot (crop cut) approach
      Can only justify a crop cut for a few important/major crops

Method
• Establish crop cut plots after the farm is planted

• Select sample of plots using a random procedure
  Problem: staff often avoid sampling areas with “no crop” because they see the crop all around. But excluding “no crop” areas introduces bias

• Number of sample plots & plot size depends on variability in field (e.g., crop planted, planting method—scatter vs. rows—monocropped vs. intercropped field, density of crop stand)
  o Some “experts” recommend a 10% crop cut; others suggest 3 plots, 50m² (rectangular, but random orientation)
  o FAO recommends:
    • For dense crop stands: 1-5 m² (rice)
    • For intermediate stands: 10-25m² (maize)
    • For dispersed crops: 100 m² (trees)
Problems with crop cuts:
- If edges of field are included, yield may be overestimated by > 20%
- Skill is required to accurately lay out rectangle, weight, calculate yield (FAO fertilizer trials in Sierra Leone)
- Logistics difficult: you should be present when farmer harvests
  - But harvest time is a function of weather, labor availability, and crop (tubers?), so you may miss the harvest
  - Alternative: ask farmer to store/save crop cut plots separately from his total harvest
- Coefficient of variation (i.e., std. deviation/mean) for crop cut estimates are often high, even if done “properly” (IRRI)

b) Record entire harvest (very accurate)

Method
- Enumerator observes harvest, records data

Problems with recording entire harvest:
- Very time consuming, since you should be present at harvest
- Requires prior arrangements, especially if fields/crops are harvested at different times
- But, it’s possibly teach farmer to record the harvest
  - Example: project asked farmers to harvest beans in a special bucket, with marks showing how full the bucket was (¼, ½, ¾) and record these data for the enumerator
4. Indirect(proxy methods

- Marketing Board records: used for cash/export crops, if all of the crop is sold to the board

- Landlord’s records, if landlord observed harvest
  (e.g., Philippines, landlord get 50% of tenant’s crop)

- Farmer estimate of value of sales, if farmer sold all of the crop
  (i.e., divide total revenue by price/kg, especially appropriate for estimating cash crop production—coffee, cacao)

5. Farmer’s own production estimates OK if:

- Farmers use standard units & weight of standard unit is known
  (e.g., record bags, sheaths harvested—Philippines—and weight a sample of the units to estimate the conversion factor)

---

D. Labor Measurement

1. Possible reasons for wanting precise estimates of labor

   - Is the most important household resource/production input, especially in traditional farming systems

   - Is a key variable in labor use (returns to labor), cost of farm production, and technology impact studies
     (e.g., earnings from farming vs. migrating to city; impact of mechanization on employment)

2. Measurement difficulties (sources of non-sampling error)

   - Farmers draws on diverse labor sources
     (e.g., family, hired, exchange labor; male, female, kids)

   - Labor stock is a poor proxy for estimating actual labor used
     (e.g., men migrate during season, do off-farm work)

   - Farmers often have only a vague concept of time, so their estimates of hours worked are unreliable
• What reporting period is best?:
  o Do you ask about typical/normal year or this past year (e.g., this past year may be an abnormal year)?
  o Choice of reference period depend on proposed analysis

• Do you only include work time or also ask farmer to estimate travel time?

• Labor efficiency varies by labor source, so you can’t sum labor use to estimate total person days (MD) of labor input
  o Problem of equivalent units, must decide if 1 MD = 1 woman day = 1 child day?
  o Some researchers convert labor data to PD equivalents (e.g., woman day=0.7 PD, 1 child day = 0.5 PD, but be cautious! Can’t always assume that men can do more work than women…it depends on the task)

• Hours in a PD aren’t equal for all tasks, so using a standard (i.e., 8 hour = 1 PD) to convert hours to PDs biases data
  o Different tasks may require different conversion factors (e.g., 1 plowing MD = 4 hours, 1 weeding MD = 8 hours)

• Difficult to allocate all labor use to a specific crop
  o Farmers in a mixed cropping system typically work on several crops during a given day
  o So, you may need to allocate some labor to a general category (i.e., “not specified” labor)

• Exchange labor is difficult to value
  o Exchange labor it typically valued at its opportunity cost, but what is opportunity cost (Ryan, ICRISAT)?
  o Also, must include cost of food given to labor
• It’s difficult for farmers to accurately estimate the amount of continuous, non-registered labor used (e.g., weeding, child labor)

• Farmers may use different technologies in different fields and/or parcels within a given field
  For example:
  o Plow with oxen vs. human labor
  o Harvest with sickle vs. “ani-ani” knife
  o So, must record & report labor by technology type

• Other problems?

3. Approaches for collecting labor data

   Best approach depends on research objectives, proposed analysis (i.e., precision needed) & resources available

   a) Single-visit survey (ask farmer to recall data for past season)
      • Potential for much measurement error, if you try to collect non-registered continuous data using a recall survey

      • Single-visit survey OK if you only want to estimate the approximate amount of labor used

      • But best to only collect data about major farming operations (e.g., land preparation, planting, harvest)
b) Cost-route survey (farm record keeping)
   - Visit farmers every 1 or 2 weeks to collect data on labor & other inputs used since your last visit
   - Short recall period reduces recall error, so data are very accurate
   - To improve accuracy, can ask farmer to mark a calendar to show days worked & type of work done each day
   - But it’s very labor intensive/costly to collect RK data (e.g., requires much staff time, 1 staff per 10–15 farmers)
   - Typically, resources are only available to record labor inputs for a small sample of farmers, which threatens the representativeness of the data

4. Labor data collection/measurement strategies
   a) Identify minimum data needed
      Example, consider only asking farmers about:
      - Most important labor inputs (LP, planting, harvesting)?
      - Critical determinants of yield (weeding)?
      - Labor input that vary most from farmer-to-farmer?
   b) Reduce recall period by scheduling data collection to coincide with farm operations
      (e.g., collect planting labor at beginning of season, weeding data at mid season, etc.– i.e., during 2-3 rounds/visits)
   c) Select an intensive data parcel (farmers’ largest parcel) for each crop/technology of interest, only collect data for this parcel & use these data to estimate labor input estimates
d) Combine cost-route data collection with a single visit survey

For example:
For total sample, collect basic economic data (e.g., area, crops grown, tenure, purchased inputs)

For small subsample of farmers, select an IDP & use precise measurement (i.e., measure area & output) and cost-route method (i.e, frequent recall surveys) to collect labor data

- Use IDP standards to generate representative budgets
- Use cost-route data for linear programming analysis
- Use small sample to estimate labor use:
  - Reduces measurement error
  - But small sample may not be representative

5. Characteristics of a complete data set for whole farm analysis

Key variables: HH, crop, operation, labor type, technology:

<table>
<thead>
<tr>
<th>Field</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
</tr>
</thead>
<tbody>
<tr>
<td>All enterprises/crops: maize, cassava, sweet potatoes, etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All activities/operations (by technology used)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Land preparation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Planting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Weed control Tech. level 1: hand weeding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tech. level 2: herbicide</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Harvesting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For each operation, labor (by type) Male-Female-Kids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hired-Family-Exchange</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cost-route gives high accuracy, but expensive and typically creates major data management problems
E. Purchased Inputs
(e.g., seed, fertilizer, fungicide, insecticide, herbicide)

1. Possible reasons for wanting precise estimates of input use
   • Required to estimate production costs/profitability

2. Measurement difficulties (sources of non-sampling error)
   • Inputs are sold in several different formulations & units
     For example:
     o Variety: same variety may have a different name in different locations
     o Fertilizer: lbs, kgs; 45-0-0, 15-15-15, foliar
     o Insecticides/fungicides: mls, ounces, gms; many brands used
     o Herbicides: liquid, granular; many brands used
   • Farmers may have difficulty recalling amount applied to each field/parcel (or may tell you amount bought, not amount applied)

3. Data Requires to Estimate Costs of Production
   • Identify/draw a rough map of farmer’s fields
   • Solicit/collect data by field/parcel, referring to the map
   • Ask farmer to fully describe each input applied (not bought), including conversion factor needed to standardize data
     For example
     o Seed: variety name, source (own harvest, neighbor, seed dealer), number of units (e.g., bags), & weight of unit
     o Fertilizer: formulation (e.g., SSP, 15-15-15, urea), number of units (e.g., bags), & weight of unit
     o Insecticide/fungicide: brand, number of units (gm, oz, ml, bottles), & weight of unit (e.g., oz/bottle)
     o Herbicide: brand, formulation (e.g., liquid or granular), number of units (liter, kg, bag, bottle), & weight of unit (kg/bag, oz/bottle)
   • Collect data on cost per unit for each input applied (e.g., pesos per bag/bottle)
4. Data Collection Strategies
   • **Single-visit** recall survey
     May be OK, IF:
     - Farmer plants only a few fields & applies few inputs
     - You only need to collect approximate data
   
   • **Multi-visit** recall survey
     (i.e., collect data during several visits, which follow planting, mid-season, harvest)
     - Produces more accurate data than single-visit surveys
     - Consider if you need to collect relatively accurate data

   • **Farm record keeping** (e.g., visit farmer ever 1-2 weeks),
     - Consider if you need very accurate data and/or farmers plants many fields, uses many inputs

F. Income and Expenditure Measurement
   (i.e, all income sources & purchases made)

   **Def.**
   *Total household income*—value of HH’s income from all sources, including agricultural production, wages, remittances received, in-kind gifts
   
   *Total household expenditures*—value of all HH purchases, remittances given, in-kind gifts

1. Possible reasons for requiring accurate measurement

   Want to assess HH’s
   - Absolute/relative socio-economic status
   - Incidence of poverty/food insecurity
     (e.g., HHs above/below poverty line)
   - Market participation: sales/purchases (i.e., net maize buyers/sellers)
   - Impact of devaluation/increase in maize price
   - Resources available to invest in agriculture
2. Measurement difficulties (non-sampling error)

Includes all problems noted for collecting accurate data for estimating crop income,

Plus:

- Respondents may have difficulty recalling
  - Small & reoccurring sources of cash income/expenditures
  - Data for a “specific” time period (i.e., year, border/telescoping—earlier/later data included)
  - Home-consumed subsistence items used (e.g., firewood, livestock, palm wine, thatch)
- Difficult to value home-consumed items (typically use “opportunity cost”)
- Respondents may falsely report sensitive income sources (e.g., income from illegal activities; fear of taxation)

3. Measurement approaches

- **Single visit** survey
  - Only reliable for big/major items
    - Accuracy depends on length of recall period (e.g., 1 year is very unrealistic!)
- **Multi-round** survey
  - Typically carried out monthly (at a minimum) for 12-15 months

4. Strategy for collecting detailed income/expenditure data

- **Steps**
  - Define the income/streams (i.e., identify all income components: crop sales, wages, remittances, in-kind gifts)
  - Create a recording form
  - Ask respondent about each **transaction type**
• Record data for each transaction as 1 record
• Collect data from each adult family member (husband, wife)

• Structure of a complete income/expenditure data record

Description of the Transaction

<table>
<thead>
<tr>
<th>HH</th>
<th>HH member</th>
<th>Type</th>
<th>Source</th>
<th>Date</th>
<th>Amount</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• Be sure to:
  o Specify key variables for each transaction: HH, transaction type, HH member, source
  o Include farm sales, non-cash exchanges, & retained farm production (harvest) transactions, convert to monetary value
  o Limit inquiry to essential big items (know the system)

---

G. Livestock-Related Values

Difficulties in collecting accurate date vary with type of production system, type of animal, & data sought

a) Difficulty depends on type of system & number of animals owned/held
  • Home kept: easiest (simply count)
  • Communally grazed: more difficult if few in number (recall OK)
  • Nomadic: very difficult to find respondent, so numeric data is typically estimated from aerial strip counts, water point counts

b) Animal type
  Accuracy of recall data depend on the economic important of the animal
  • Cattle: easy
  • Sheep/goats: less accurate
  • Poultry: very difficult (important?)
c) Types of data

- Number of animals: easily recalled, but must specify situation (e.g., animals owned, cared for)
- Age: can count teeth (i.e., cattle)
- Technical coefficients (i.e., productivity characteristics)
  - Births (offspring/year), recall OK
  - Deaths, recall OK, but farmers may exclude stillborn
  - Birthing intervals (i.e., time between each kidding, calving), recall OK
  - Off-take/sales data, recall OK
  - Weight gains (must measure before/after)

c) Difficulty of collecting animal technical/productivity data exaggerated!

- Indonesia: RA vs. monitoring (little difference)

e) But must:

- Assess/confirm accuracy of farmers' estimates in local setting
- Get assistant from local animal scientist to convert technical jargon into local equivalent
- Collect data animal-by-animal (name?)
- Specify time frame for data requested (e.g., season)
H. Inventory Values (i.e., wealth indicators)

1. Building Value
   Problems
   • Original cost gives low estimate, if building is old
   • Replacement cost is hard to obtain (e.g., large labor input)
   Options
   • Use square footage as proxy (plus description of type of structure)
   • Ask key informant to estimate value, based on value of comparable structure with known cost

2. Equipment Value
   Problems
   • Same as above

   Options
   • Only collect data on type of equipment owned & condition of equipment (i.e., working vs. not working/broken)
   • Consider only “important” items (e.g., plow, but not sickle)

3. Draft animals/livestock value
   Problem
   • All livestock not comparable (e.g., trained oxen more valuable than untrained oxen)
   • Inventories fluctuate throughout the year
   Options
   • Only collect data on number of animals by age class, sex, type (e.g., draft animals, beef cattle, milk cattle)
   • If you need to estimate value:
     o Check market for price of “comparable” age, sex, type of animal
     o Assign “average” value by type
4. Value of stored grain

**Problem**
- Price is lowest at harvest, increases over time
- But, post-harvest price increase reflects returns to storage (e.g., income generated by storage activity, not production)

**Options**
- If pan-seasonal pricing exists,
  - Value stored grain at harvest price
- If seasonal price varies over time (most common situation)
  - Value stored grain at current market price (i.e., opportunity cost)

I. Price/value of harvested grain (and sales)

**Problems**
- Must value harvest to estimate income from crop production
- However, respondents in one village may report a very different price received for the same commodity
- But since commodity prices can vary, depending on buyer/point of sale, farmers’ price may be OK?

**Options**
- Collect data (i.e., amount sold, price received, buyer, location of sale) 1-3 mos. after harvest (most sale made soon after harvest);
- Use “average” price farmer in village received in a given month, if you are pretty sure that the price variation is due to recall error
- Must vary average price, if grain was sold to different buyers and/or in different markets
- Middlemen may have “local time series” data that you can use as a cross check on prices farmers’ reported
J. Training and Instructions to Enumerator Are Critical to Minimize Non-Sampling Error

• Must
  • Teach enumerators what data needs to be collected and how it should be recorded during pre-survey training
  • Include instructions in the questionnaire that identify the specific type of data that the enumerators should collect
  • Include definitions in the questionnaire to specify meaning of terms like parcel, owned, etc.
  • Keep in mind, the devil is in the details!!!