The Process of Poverty-Scoring Analysis

Mark Schreiner

22 October 2014

Abstract

This report is a high-level outline of the main steps and concepts in the process of analyzing estimates based on poverty scoring. It aims to help managers in local, pro-poor organizations to guide the process wisely through all stages so that poverty scoring can usefully inform questions related with social performance.

Acknowledgements

Author

Mark Schreiner directs Microfinance Risk Management, L.L.C. He is also a Senior Scholar at the Center for Social Development at Washington University in Saint Louis.

The Process of Poverty-Scoring Analysis

Introduction

Poverty scoring uses responses to ten verifiable, inexpensive-to-collect questions to estimate the likelihood that a household has consumption below a poverty line.[1] The average of the poverty likelihoods of households in a group is an estimate their poverty rate at a point in time. If a scorecard is applied twice—whether to all clients in a single sample or to different clients in two samples—then the change in poverty likelihoods is an estimate of the change in the poverty rate through time.

Poverty scoring is both relatively simple (compared with alternatives with similar goals) and absolutely complex (involving not only mathematical details but also management challenges). In particular, managers must lead a preparatory process that identifies questions, defines a relevant population, picks an analytical approach, draws a representative sample, chooses non-scoring data items to collect, interviews households, and records responses. Managers then lead an analytical process that sets up data, derives scoring estimates and juxtaposes them with non-scoring data, analyzes the results for clues for possible causes and promising areas for further inquiry, and clearly reports the estimates as well as speculates about cause-and-effect relationships that might explain the numbers.

This report is a high-level guide to this process for managers in pro-poor organizations who seek to use poverty scoring so as to improve the information available for social-performance decisions. It outlines a good-practice approach (Figure 1)[2] whose 14 steps fall into four groups:
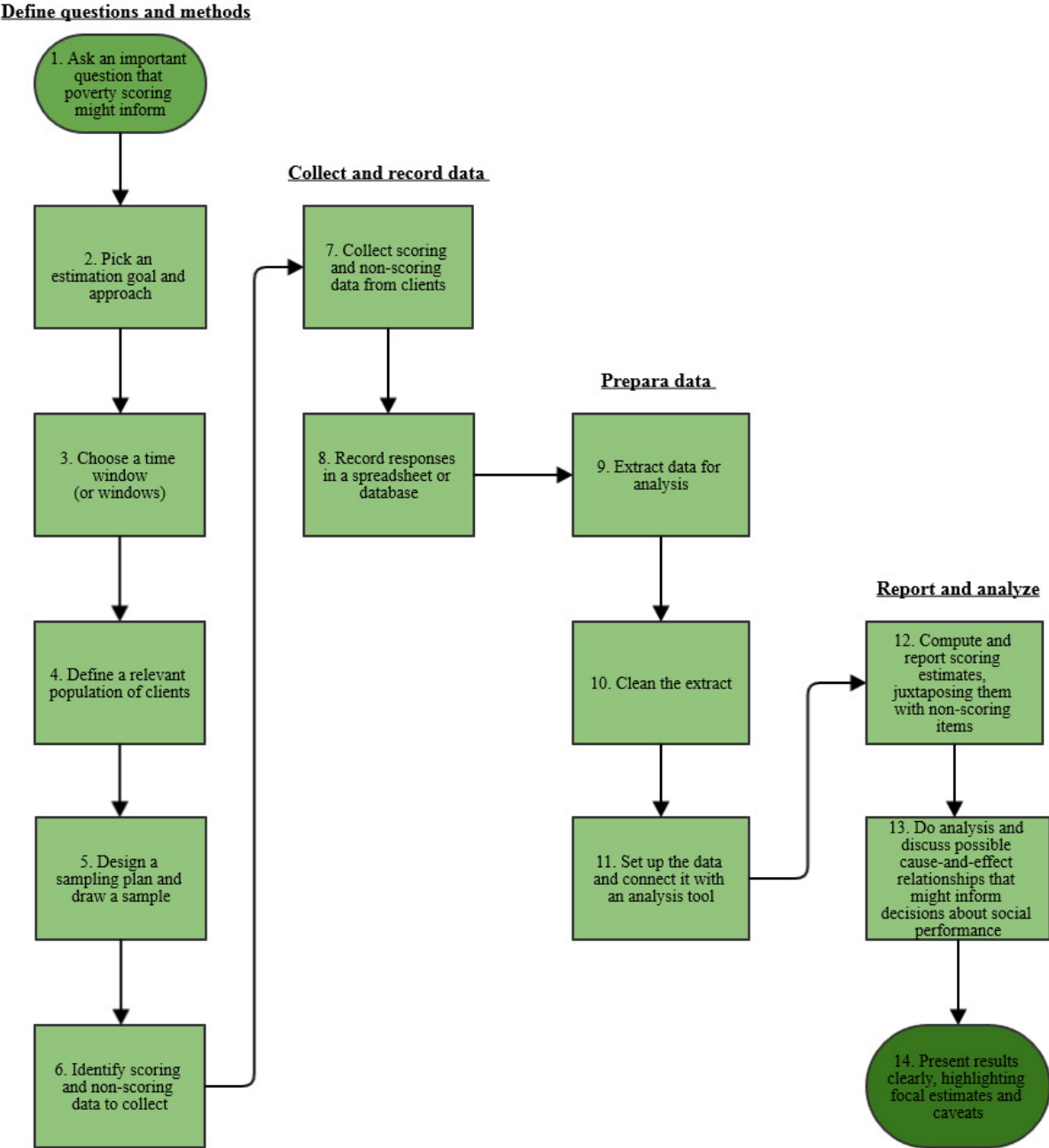
- Define questions and methods
- Collect and record data
- Derive scoring estimates
- Analyze and report results

The rest of this report discusses each of the above steps in detail.

---

[1] There are Simple Poverty Scorecard®-brand poverty-assessment tools for more than 60 countries (`SimplePovertyScorecard.com`), and they are used by more than 200 pro-poor organizations (Grameen Foundation, 2014). They are the same as what GF calls the Progress out of Poverty Index®. The PPI® is a performance-management tool that GF promotes to help organizations achieve their social objectives more effectively.

[2] Parts of this process were tested in an exercise that reviewed (and improved) poverty-scoring analysis done by Vision Fund/Ecuador.

Figure 1: Poverty-Scoring Analysis in 14 steps

1. Ask an important question that poverty scoring might inform

2. Pick an estimation goal and approach

3. Choose a time window (or windows)

4. Define a relevant population of clients

5. Design a sampling plan and draw a sample

6. Identify scoring and non-scoring data to collect

Collect and record data

7. Collect scoring and non-scoring data from clients

8. Record responses in a spreadsheet or database

Prepara data

9. Extract data for analysis

10. Clean the extract

11. Set up the data and connect it with an analysis tool

Report and analyze

12. Compute and report scoring estimates, juxtaposing them with non-scoring items

13. Do analysis and discuss possible cause-and-effect relationships that might inform decisions about social performance

14. Present results clearly, highlighting focal estimates and caveats

<u>Define questions and methods</u>

1. Ask a business question whose answer matters

The purpose of using poverty scoring is not to use poverty scoring; it is to inform business questions that matter, especially for social performance. For most questions facing a pro-poor organization, poverty scoring is not relevant, but it may help inform questions related with the depth and breadth of poverty outreach (Schreiner, 2002).

In particular, an existential question for pro-poor organizations—and their supporters—is the extent they reach the poor and whether their clients' poverty is decreasing through time. Poverty scoring can inform this question. In fact, the main impact of poverty scoring so far has been to disabuse many organizations of the belief/claim that almost all of their new clients are poor. This has prompted some pro-poor organizations to be more explicit and intentional in their efforts to reach poor clients, whether by seeking reasons for their lower-than-expected poverty outreach, by setting poverty-related goals for field agents or field offices, by opening field offices in poorer locations, by targeting poorer clients, or by developing products/services that are more attractive to poorer people.

2. Pick an estimation goal and approach

The two main uses[3] of poverty scoring are to estimate:

- Poverty rates in a window in time
- Changes in poverty rates through time

A poverty rate in a window in time is estimated as the average poverty likelihood of clients sampled in the window.

Changes poverty rates between two time windows are estimated by:

- Scoring two independent samples at two points in time, taking change as the difference between follow-up and baseline point-in-time estimates, and dividing the change by the average years between baseline and follow-up interviews, or
- Scoring each client in a single sample at least twice and esimating the rate of change as the weighted average of each client's change in the poverty likelihood, where each client's weight is the years between his/her first and last scores

---

[3] The use of scoring for segmenting clients for targeted services is not discussed here.

Figure 2 uses hypothetical data to illustrate the estimation of a poverty rate in a single time window (in this case, June to December of 2012). The estimate is the average of the poverty likelihoods of clients scored in the window, that is, (45.3 + 32.1 + 66.7 + 5.0 + 10.2 + 32.1 + 2.1 + 43.2 + 97.2) ÷ 9 = 333.9 ÷ 9 = 37.1 percent.

Figure 2: Estimating a poverty rate at a point in time

Client	2011							2012												2013		
	0 Jun	1 Jul	2 Aug	3 Sep	4 Oct	5 Nov	6 Dec	7 Jan	8 Feb	9 Mar	10 Apr	11 May	12 Jun	13 Jul	14 Aug	15 Sep	16 Oct	17 Nov	18 Dec	19 Jan	20 Feb	21 Mar
A	45.3															39.8				35.8		
B		32.1																				
C			66.7														60.2					58.8
D				5.0														22.3				
E				10.2																		
F					32.1					43.2						30.0						
G						43.2				22.2												
H							97.2					33.3									44.3	
I																						
J													23.2								18.2	
K														50.0								
L				2.1												4.0						10.0

Cells are the poverty likelihood (%) derived from scoring in months when a given client was scored.
This example assumes interviews are on the first of the month, months have 30 days, and years have 360 days.
Months are numbered consecutively starting with 0 for June 2011 to 21 for March 2013.

Figure 3 shows how to use two independent samples at two points in time to estimate change through time. It is the difference between the average poverty likelihood in the first time window (June to December 2012, 37.1 percent) versus the second time window of January to December 2013, that is, (43.2 + 22.2 + 33.3 + 23.2 + 50.0 + 39.8 + 30.0 + 4.0 + 60.2 + 22.3) ÷ 10 = 328.2 ÷ 10 = 32.8 percent.

The change between baseline and follow-up is 37.1 – 32.8 = 4.3 percentage points. But how much time passed for this change to occur? The estimate is most meaningful when expressed as an *annual rate of change*. This is the difference between the follow-up and baseline estimates, divided by the difference between the average date of interviews for clients in the first window and the average date of interviews for clients in the second window. The average first-window interview is in month (1 + 2 + 3 + 4·4 + 5 + 6) ÷ 9 = 33 ÷ 9 = 3.7), while the average second-window interview is in month (9·2 + 11 + 12 + 13 + 15·3 +16 + 17) ÷ 10 = 132 ÷ 10 = 13.2). This implies an annual rate of change (reduction) in poverty of $\dfrac{37.1 - 32.8}{\left(\dfrac{13.2 - 3.7}{12}\right)} = \dfrac{4.3}{0.8} = 5.4$ percentage points per year.

Figure 3: Estimating an annual rate of change in poverty through time with two independent samples in two time windows

Year/month scored

Client	2011							2012												2013		
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
A		45.3														39.8				35.8		
B			32.1																			
C				66.7													60.2					58.8
D					5.0													22.3				
E					10.2																	
F					32.1					43.2						30.0						
G						43.2				22.2												
H							97.2					33.3								44.3		
I																						
J													23.2								18.2	
K												50.0										
L					2.1											4.0						10.0

Cells are the poverty likelihood (%) derived from scoring in months when a given client was scored.
This example assumes interviews are on the first of the month, months have 30 days, and years have 360 days.
Months are numbered consecutively starting with 0 for June 2011 to 21 for March 2013.

Figure 4 and Table 1 show how to estimate change through time with one sample in one time frame in which all clients are scored at least twice. For the six clients scored twice in the window from January 2012 to March 2013, the annual rate of change (reduction) in poverty is the weighted average of the client-by-client changes in poverty likelihoods, where the weights are the years between each client's first and last scores. This turns out to be a reduction of $1.5 \div 3.1 = 0.5$ percentage points per year (Table 1). The full calculation in a single expression is:

$$\frac{0.33 \cdot 4.0 + 0.42 \cdot 1.4 + 0.50 \cdot 13.2 + 0.67 \cdot (-11.0) + 0.67 \cdot 5.0 + 0.50 \cdot (-6.0)}{0.33 + 0.42 + 0.50 + 0.67 + 0.67 + 0.50} = \frac{1.5}{3.1} = 0.5.$$

Table 1: Year-weighted average of client-by-client poverty likelihood

Client	Poverty likelihood			Month scored			(Change in pov. like.) x (Change in years)
	First	Last	Change	First	Last	Change in years	
A	39.8	35.8	4.0	15	19	4÷12 = 0.33	1.32
C	60.2	58.8	1.4	16	21	5÷12 = 0.42	0.59
F	43.2	30.0	13.2	9	15	6÷12 = 0.50	6.6
H	33.3	44.3	−11.0	11	19	8÷12 = 0.67	−7.37
J	23.2	18.2	5.0	12	20	8÷12 = 0.67	3.35
L	4.0	10.0	−6.0	15	21	6÷12 = 0.50	−3.00
					Sum:	3.1	1.5

The year-weighted average change (reduction) in poverty rate is $1.5 \div 3.1 = 0.5$ percentage points per year. That is, poverty decreased by 0.5 percentage points/year.

Figure 4: Estimating an annual rate of change in poverty through time with one sample in one window in which all clients are scored at least twice

Client	Year/month scored																					
	2011							**2012**												**2013**		
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
A		45.3														39.8					35.8	
B			32.1																			
C				66.7													60.2					58.8
D					5.0													22.3				
E					10.2																	
F					32.1						43.2					30.0						
G						43.2				22.2												
H							97.2				33.3								44.3			
I																						
J										23.2									18.2			
K												50.0										
L					2.1											4.0						10.0

Cells are the poverty likelihood (%) derived from scoring in months when a given client was scored.

This example assumes interviews are on the first of the month, months have 30 days, and years have 360 days.

Months are numbered consecutively starting with 0 for June 2011 to 21 for March 2013.

Given their assumptions, all three approaches give unbiased estimates. Of course, the two ways of estimating of the rate of change through time use different data and so give different estimates.

3. Choose a time window

Given a business question and a measurement approach, the analyst then chooses a time window (or two windows) within which a poverty rate (or between which a change in a poverty rate) is measured. Because it takes time to interview a sample of clients, a typical time window is not measured in days or week but rather in months or years.[4]

Figure 5 on the next page is a flowchart that depicts the three approaches to finding poverty-scoring estimates. After defining the question and choosing an analysis approach, the time window/windows is/are set.

For estimating a poverty rate for a population at a point in time, the analyst defines a single time window, and then—as discussed earlier—averages the poverty likelihoods of all clients scored in that window.

Estimating the change in a poverty rate through time requires defining:

- Two time windows (if two different samples from one population are scored)
- One time window (if each client in a single sample is scored more than once)

With two time windows and two independent samples, an organization measures change by drawing one sample from a given population—such as all current clients—in one time window (say, 2010 to 2011) and then a second sample from (what is assumed to be) that same population in a second time window (say, 2013).

This case with two time windows also applies if an organization samples continuously, for example, if a microlender applies scoring to a random sample of 10 percent of all clients (new or old) who apply for a loan. With continuous sampling, the two time windows are defined after-the-fact and may or may not cover the entire period in which sampling took place. This two-window approach can also be used with a continuous census in which all clients are repeatedly scored (say, with each new loan or every three years).

With two independent samples from the same population, the estimate of change through time is the difference in households' average poverty likelihood between the baseline sample (first window) and the follow-up sample (second window), divided by the number of years between the average baseline interview date and the average follow-up interview date.

With a single time window (say, from 2010 to 2013) and each client scored at least twice, change could be measured if scoring happens continuously (say, every time a sampled client applies for a loan) or if scoring is done only in sub-windows within the overall time window (say, in the first half of 2010, and then in the second half of 2013). This single-window approach also works with a continuous census.

---

[4] A "point-in-time" estimate actually covers a period of time.

Figure 5: Flowchart for two broad approaches to poverty-scoring estimates, with time windows
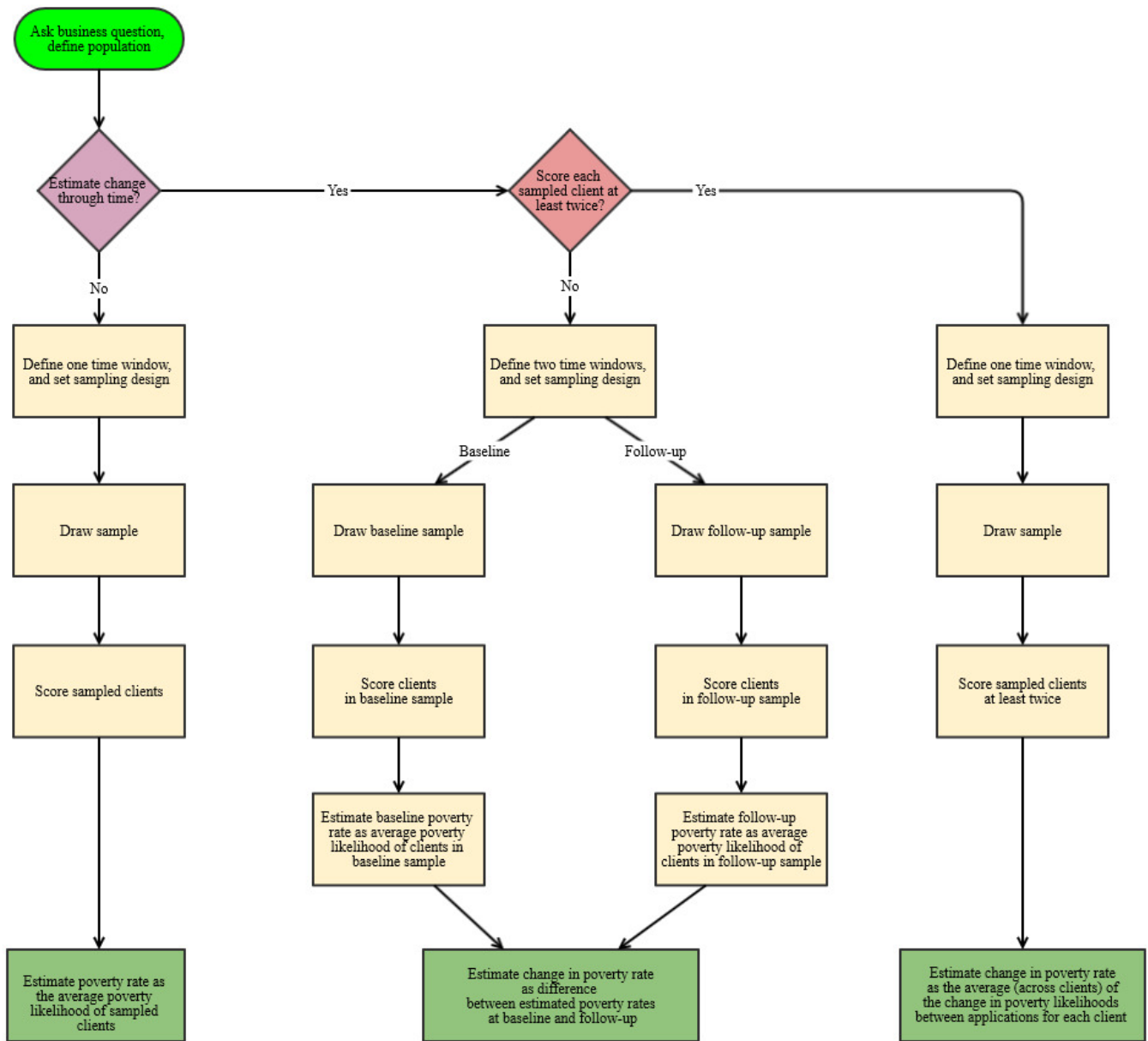
The two approaches to measuring change through time give similar—but not identical—estimates. Which is to be desired depends on the level of accuracy desired, the cost of applying the scorecard, and whether it makes more sense to assume that the population does not change over time (two windows) or to define the population to be only those clients who do not exit over time (one window).

All else constant, the one-window approach (scoring each client at least twice) is more accurate. Also, the one-window approach is probably less costly: it is more difficult to draw two (mostly non-overlapping) samples and then visit each client than to draw a single sample and then track down the same clients twice. Nevertheless, an organization's business question may pertain to all clientele in a population (as with two windows) rather than only to clients who do not exit (as with one window).

4. Define the relevant population

Scoring estimates are based on samples of clients.[5] The sample is representative of a well-defined population for which any given client's membership is set by a few clear rules. *Representative* means that—on average in repeated samples—the distribution of characteristics in the sample match the distribution in the population. Estimates from a representative sample can be extrapolated to the population. The two most common ways to get a representative sample is random sampling and census.

For estimating change in poverty rates through time, the two-window approach defines the population assuming that there are no drop-outs nor kick-outs or that drop-outs and kick-outs occur at random.

In contrast, the one-window approach defines the population to exclude any drop-outs or kick-outs. This maintains an unchanged population through time, but it probably makes the population differ from the clientele as a whole at any point in time.

5. Design a sampling plan, and draw a sample

The simplest sampling plan is a *census* that scores everyone in the population. While it is more costly to score all clients rather than some clients, a census has lower costs for figuring out who is in the sample and for communicating that to field agents. A census also maximizes flexibility for later analyses and produces the most precise estimates. If a population is small (say, 500 clients or less), then a sample would cost almost as much as a census anyway.

With larger populations, sampling reduces the number of scoring applications and thus may reduce costs, possibly compensating for its managerial complications.

In a *simple random sample*, each client has the same probability of being selected. For example, if the population is 10,000 clients, and 1,000 are sampled, then each client has a sampling probability of $1,000 \div 10,000 = 0.10$, and each sampled client represents $1 \div 0.10 = 10$ clients in the population. This is the client's *sampling weight*.

---

[5] A *census* is a special case of a sample that includes an entire population.

In a simple random sample—as in a census—all sampled clients have the same sampling weight.

The sampling design should be documented, along with each client's sampling probability and sampling weight. A sampled client's sampling weight is 1 (one) divided by her sampling probability. The sampling weights are used to weight averages when deriving estimates. This maintains representativity and thus allows extrapolation from the sample to the population.

Simple random samples are costly in that they spread the sample across all field offices and field agents, so all agents must be trained to apply the scorecard and all offices must process completed scorecards. To reduce costs, an organization could cluster the sample by field office (only clients in a sample of offices are sampled) and/or by field agent (only clients in a sample of agents are sampled).

For most organizations, it is wise to start with a simple design and to estimate a poverty rate at a point in time. If an organization uses these basic estimates to inform decisions, then it might make sense to entertain additional complications.

An organization can set the overall sample size n using formulas documented with each country's scorecard. In practice, sample size is usually driven by cost considerations. Rather than worry about whether a sample size is statistically optimal, it is better to recognize that sample size depends mostly on cost and that some sizes (such as 1,000 or 10,000) are disproportionately likely to seem "big enough". The key is not "statistical significance" but rather to have a representative sample of a well-defined population that is relevant for an important business question.

6. Identify scoring and non-scoring data items to collect for later juxtaposition

Analysis puts poverty-scoring estimates in context to suggest causes, feasible ways to influence causes, and potentially fruitful areas for additional investigation.

Putting poverty-scoring estimates in context to derive information useful for decision-making requires data beyond that of the scorecard. This section suggests non-scoring items to collect. Of course, the actual items collected will depend on the business question(s), the specific context, and costs. The goal here is to help analysts to think about what would be useful and to avoid overlooking central items.

Scoring data items:
Items marked with an asterisk must be recorded.

Record key:
Each interview is stored as a record (in a database) or a row (in a spreadsheet) that is uniquely identified by the combination of three fields:

- Client identifier*, and
- Scorecard identifier*, and
- Date of scorecard application*

Additional house-keeping identifiers:
- Sequential interview identifier*
- Sample identifier*
- Selection probability*
- Sampling weight* (1 ÷ selection probability)
- Cluster identifier
- Stratum identifier
- Enumerator identifier

Client-related identifiers:
- Client identifier*
- Number of household members*
- Date the client first participated with the organization
- Identifier of the client's field office
- Identifier of the client's field agent
- Other organizational identifiers
- Client contact information
- Other items (*e.g.*, the client's preferred language, or the length of the interview)

Responses to scorecard indicators:
- *Response to the first indicator*:[*] Record the character code of the response (such as A, B, or C), not the text description (such as "None", "Television only", or "Television, and DVD or VCR"), nor a numeric code (such as 1, 2, or 3), nor the number of points (such as 0, 4, or 7)
- *Response to the second indicator*[*]
- . . .
- *Response to the last indicator*[*]
- *Score.*[*] This is the sum of points in a scorecard application
- *First poverty likelihood*[*] (as documented for a given country's scorecard)
- *Second poverty likelihood*[*]
- . . .
- *Last poverty likelihood*[*]

Non-scoring data items:

Juxtaposing poverty likelihoods with non-scoring data on clients helps put scoring estimates in context. This is the essence of analysis.

Non-scoring items are not required, but there is little to analyze without them. An organization may already collect and record some non-scoring items, say, in a loan application or a biannual client survey. If other items are expected to be useful, then they could be collected from now on, perhaps in tandem with the scoring interview.

The list is below not definitive nor exhaustive. Rather, it suggests broad categories to spark thought about what data might usefully be crossed with scoring estimates.

Client characteristics:
- Date of birth
- Male/female
- Marital status
- Urban/rural location
- Client status (active/drop-out/kick-out/other)
- Race/ethnicity/caste/language (social categories based on easily observable characteristics fixed at birth)
- Highest level of education completed by the client
- Employment:
    — Employment status of client (Self-employed in agriculture, self-employed in non-agriculture, not working, unpaid worker, day laborer, wage employee, or salary employee)
    — Whether any household members are day laborers
    — Whether any household members work in agriculture
    — Whether any household members have a permanent wage or salary job

- Finances:
    - Level of free cash flow (if already collected)
    - Whether the client's household has any formal savings accounts
    - Whether the client keeps written records of cash received and cash spent
- Other direct indicators of poverty:
    - Receipt of means-tested transfers
    - Field agent's subjective perception (poor/average/well-off)
    - Others

Characteristics of product/service used:
- Type of service
- Delivery method
- Characteristics of loan products:
    - Amount disbursed
    - Number of installments
    - Amount of installment
    - Price (fees and interest per dollar outstanding per month)
    - Type of guarantee/collateral
    - Use of the loan (as supposed by the lender)
- Aspects of a client's use of a loan product:
    - Whether the client previously received loans from another formal lender
    - Whether client has a record in the credit bureau
    - Number of previous loans received by the client from the organization
    - Use of the loan (as stated by the lender or averred by the client)
    - Aspects of repayment performance in the most-recent paid-off loan:
        - Days of longest spell in arrears
        - Percentage of installments with any arrears
        - Average days in arrears per installment
- Characteristics of saving services:
    - Minimum opening deposit
    - Fees (monthly and per-transaction)
    - Interest rates (or other incentives for savers)
    - Limits on withdrawals
- Aspects of a client's use of saving services:
    - Current balance
    - Average balance in past year
    - Number of transactions in past year
    - Change in balance in past year

An organization should analyze poverty first in relation to data that it already collects and uses for decision-making, as revealed by a careful scan of client in-take forms, loan applications (for microlenders), databases, and existing management reports/analyses. The next step is to juxtapose scoring estimates with the most promising non-scoring items. If that analysis is used to inform managerial decisions, then the analyst can identify other promising indicators to start to collect from now on.

Collect and record data

7. Apply the scorecard to clients

This section outlines implementation advice, as high-quality analysis requires high-quality data and thus high-quality implementation.

Do a pilot. Toohig (2008) and IRIS Center (2007) offer detailed guidance. Organizations using poverty scoring should study both of these documents carefully.

Interview the client at home. The scorecard interview should mimic the national survey that provides the data upon which the scorecard is based.

Fit scoring smoothly into operations. Combine the interview with a trip to the client's home that a field agent would do anyway.

Collect key non-scoring items. Collect scoring items before non-scoring items.

Follow the sampling plan. Enumerators must interview the households of the sampled clients, even if finding them at their homes is difficult. The respondent does not need to be the client; a knowledgable adult member of the household will suffice.

Audit to keep enumerators honest. Beyond exhortations, provide carrots and sticks to ensure that enumerators do their job well.

Train enumerators well. Besides aspects of survey operations, training is mostly of a thorough review of the "Guidelines for the Interpretation of Scorecard Indicators". Enumerators should be taught to use their best judgment when the "Guidelines" do not address an issue. Careful training also shows enumerators that data quality matters.

Translate the scorecard and "Guidelines". Do not skimp; good translation improves the quality of each interview, and often scoring is impossible without it.

Ask enumerators for feedback. Many problems in the field can be caught and corrected by meeting with a few small groups of enumerators after a couple of days.

Spot-check data entry. While it is overkill to enter all data twice, random auditing makes sense. Like enumerators, data-entry operators may have sloppy habits because—before the advent of poverty scoring—the data was not used.

Do a mock analysis. After 500 or 1,000 interviews, do a dry-run of the analysis process. This checks whether all needed items are collected and provides preliminary estimates. It is better to catch a problem early than after field work is over.

Track who does not respond. Identifying non-responders (and their reasons) allows checking whether non-response has damaged the sample's representativeness.

8. Record responses in a spreadsheet or database

Analysis requires storing interview data in a spreadsheet or database.

One approach is to enter data in a home-made spreadsheet, write formulas to convert responses to points, add up points to get scores, convert scores to poverty likelihoods, and derive scoring estimates based on poverty likelihoods. A home-made spreadsheet is quick—and thus good in the proof-of-concept stage—but error-prone.

Another approach is a canned spreadsheet tool that facilitates data entry, builds a database, and computes poverty estimates automatically. A canned tool is quick and good for proof-of-concept, and it less error-prone than a home-made spreadsheet. Existing canned tools,[6] however, do not record all required scoring items, they are limited in terms of non-scoring data items, and they do not help much for measuring change through time.

After proof-of-concept, the ideal approach is integration in a full-fledged database within an organization's management-information system (MIS). This allows:

- Efficient recording of responses by an organization's regular data-entry operators
- Flexible data manipulation using standard (*e.g.*, SQL) approaches
- Centralization and automation of maintenance tasks such as back-ups
- Linking scoring items with other, non-scoring items in the MIS

Most important, database integration institutionalizes and routinizes scoring so that it is not longer seen by employees as extra work for a special project.

While MIS integration is ideal, it is also costly. Before doing it, an organization should do a proof-of-concept project to be sure that it values scoring and will use it. This allows scoring to be tested without having to wait for MIS integration. Some organizations will prefer to use a spreadsheet tool, whether canned or home-made.

---

[6] See Microfinance Risk Management, L.L.C. (*e.g.*, `microfinance.com/English/Papers/Scoring_Poverty_Data_Entry_Ecuador_EN_2005.xls`) and Grameen Foundation (`progressoutofpoverty.org/node/1674/download`).

Prepare data

9. Extract data for analysis

Prior to analysis, data must be extracted, cleaned, and connected with a tool that calculates scoring estimates and juxtaposes them with non-scoring items.

Scoring and non-scoring items can come from any or all of the three sources in the previous section. The extract should exclude:

- Clients whose households have not been scored
- Scored clients who are not in the population relevant for a given analysis
- Interviews with null data for any:
    — Element of the record key
    — Response to a scorecard indicator (and therefore a null score)

To snap into an analysis tool, the data extract must have a specific structure and content. The structure is one record (row) for each scorecard interview with a household of a sampled client in a given population, with one field (column) for each data item. The required content is the scoring items marked with an asterisk in Section 6 above. The non-required, non-scoring items are unique to each organization.

Cross-tabs are simple, understandable, and informative, so the main quantitative analysis approach is cross-tabs of categorical non-required items juxtaposed against scoring estimates. This means that the values of non-scoring items are symbols standing for codes or categories (for example, "MZQ" as a client identifier, or "0" for Male and "1" for Female) rather than true numbers that stand for a count (such 17 dollars or 3.14 meters). The symbols could be letters or integer digits. In the data extract, true counts (such as a savings balance) should be categorized by ranges (such as "A" for $1.00 to $99.99, "B" for $100.00 to $199.99, and "C" for $200.00 or more). Non-required items may take null values.

10. Clean the data extract

The data extract is cleaned before being plugged into an analysis tool. *Cleaning* is not replacing values that are somehow known to be incorrect with other values that are somehow known to be correct.[7] Rather, it is making sure that scoring data is complete, that all sampled clients—and no one else—are included, that keys are unique, that missing values[8] are coded properly, and that the cleaning process is replicable.

Deleting records with incomplete scoring data was covered in the last section.

Making sure that the extract includes all sampled clients (and no one else) means defining the population in terms of items in the extract and then deleting records whose values for these items are null or do not meet the population criteria.

Making sure that keys are unique means deleting all but one record from a set with identical keys (after figuring out which record to keep).

To make sure that missing values are coded properly, the analyst should check for blank cells in the spreadsheet, columns with unexpectedly large numbers of zeros, and undefined codes such as 99 or –1. The analyst should also ask data-entry operators and database managers how they record missing values. If a code that stands for a true, known value is also used to mark missing values, then all instances of that code should be set to null.

Responses marked "not applicable" can usually be interpreted as a meaningful category. For example, if the indicator is "Can the female head/spouse read and write?" but the household has no female head/spouse, then the response is not "not applicable" nor missing but rather "No female head/spouse".

Data cleaning is replicable when done "by program", not "by hand".[9] Changing values and deleting records by hand in a spreadsheet may seem best at first, but it is not easily un-done and re-done if mistakes are found or rules are changed.

Cleaning data "by program" means using writing software to do checks, replacements, and deletions. With software, fixing a mistake or changing a rule in (say) step 3 of 20 means fixing the code for step 3 and re-running all 20 steps of the program on a fresh copy of the original data extract.

If cleaning is done by hand, then all edits, rules applied, and affected records (identified by their keys) should be written down. This tedious investment is repaid the first time that the analyst discovers an error and must start over from scratch.

---

[7] Values known (or merely suspected) to be incorrect should be set to null.

[8] Data is missing if it exists but is not recorded. For example, all clients have a date of birth, but if that date is not known to the organization, then the corresponding item should be null, not (say) 01/01/01. This avoids confusion between cases where 01/01/01 is an actual birthdate and where it is a marker for "true value unknown".

[9] The original data extract is saved and set aside, with all cleaning applied to a copy.

11. Connect the data extract to an analysis tool

An analysis tool takes cleaned data as input, derives scoring estimates, crosses estimates with non-scoring items, and makes tables for reporting and analysis. While the analyst must still define questions and methods, collect and record data, set up data, provide meta-data, and analyze results, the analysis tool is helpful because calculating scoring estimates is tedious and error-prone. In particular:

- Estimates have three possible types (one point-in-time, two change-through time)
- Analysts may not know which estimates and cross-tabs they want
- Analysts may not be ready to compute estimates and make clear tables on their own
- Managers may want estimates both for clients and for people in client households
- Sampling weights may vary across clients
- Clients may be scored more than once in a time window
- Time windows may vary in their start and end dates
- Scorecards may vary in their number of indicators
- The number of poverty lines to which scores are calibrated may vary
- Raw estimates must be adjusted to account for known bias
- Confidence intervals have complex formulas
- Non-scoring items (and the categories of each non-scoring item) may vary in number and definition

To connect the data extract to the analysis tool, the analyst must:

- Supply the required scoring items (and ideally some non-required non-scoring items)
- Extract, clean, and structure the data extract as described in previous sections
- Specify the number of records (rows) in the data extract
- Specify a desired scoring estimate, poverty line, confidence level, and time window(s)
- For each scoring item, specify its location (column) in the data extract
- For each non-scoring item, specify its:
    — Location in the data extract
    — Number of categories
    — Code of each category
    — Text label for each category

The tool itself guides the analyst through a set of choices that specify the meta-data that maps the extract to the structure used by the tool.

The analysis tool outlined here is under development. When finished, it will have a detailed manual and will automate three major stages of poverty-scoring analysis:

- Drawing a random sample
- Data entry and storage
- Deriving estimates that are clearly juxtaposed with non-scoring items

The analysis tool lets the analyst focus first on getting high-quality data that is representative of a population relevant for an important business question and second on analyzing the meaning of the results, without having to deal with the technicalities of deriving estimates or presenting clear cross-tabs.

Report and analyze

12. Compute and report scoring estimates, juxtaposing them with non-scoring items

<u>Point-in-time estimates</u>

The basic scoring estimate is the poverty rate at a point in time. The analysis tool can derive this and present it alone as well as cross-tabbed with non-scoring items.

Table 2: Example presentation of estimate of poverty rate at a point in time

| Segment | Poverty rate (%) | | Num. poor in pop. | | 80-percent confidence interval, pov. rate (%) | | | |
| | | | | | Clients | | Persons | |
	Clients	Persons	Clients	Persons	Lower	Upper	Lower	Upper
Global:	26.0	33.9	14.3	72.4	23.7	28.3	32.2	35.5

Poverty line: 100% of national

Sample size (n) is 10 clients and 33 people.

Population size (N) is 55 clients and 214 people.

Table 2 is derived from the made-up data extract in Table 3.[10] Its point-in-time estimates for the entire sample are the focal analysis (the one presented first, with the greatest emphasis) in most poverty-scoring reports. It shows:

- Estimated poverty rates at a point in time:
  — 26.0 percent of clients
  — 33.9 percent of people who live in households with clients

The client-level rate is relevant because an organization deals directly with clients. The person-level rate is relevant because all people in a household have the same expenditure-based poverty status, and a pro-poor organization cares about the well-being of all the people in its clients' households.

Table 2 also shows:

- Number of poor in the population:
  — 14.3 clients
  — 72.4 people who live in households with clients

---

[10] This differs from the made-up data in earlier examples.

Table 3: Made-up data used in example presentations of scoring estimates

	Record key			Other scoring fields					Responses to scoring indicators												Pov. like.		Non-required items		
Client identifier	Scorecard identifier	Date of interview	Interview identifier	Sample identifier	Selection probability	Sampling wgt. (clients)	# HH members	Indicator 1	Indicator 2	Indicator 3	Indicator 4	Indicator 5	Indicator 6	Indicator 7	Indicator 8	Indicator 9	Indicator 10	Score	Poverty likelihood 1	Poverty likelihood 2	Loan type	Sex	Urban/rural
A	1	31-Dec-2009	1	1	0.30	3.33	3	C	B	A	C	A	B	A	B	C	B	42	42.0	12.1	Solidarity g	Female	Urban
A	1	10-Jun-2011	8	1	0.30	3.33	4	D	C	B	D	B	B	A	B	C	B	65	0.9	0.0	Solidarity g	Female	Urban
A	1	30-Sep-2013	23	1	0.30	3.33	4	E	C	C	D	B	B	A	B	B	B	72	0.0	0.0	Solidarity g	Female	Urban
B	1	27-Aug-2010	3	1	0.25	4.00	3	C	C	B	B	B	B	A	B	B	A	47	34.0	6.8	Solidarity g	Female	Urban
B	1	23-Jun-2011	11	1	0.25	4.00	3	C	C	C	B	B	B	A	B	B	B	55	10.6	1.2	Community	Female	Rural
B	1	30-Nov-2013	28	1	0.25	4.00	4	C	C	C	B	B	B	A	A	B	B	49	34.0	6.8	Solidarity g	Female	Rural
C	1	10-Jun-2011	9	1	0.20	5.00	5	B	C	B	B	B	B	A	A	A	A	30	71.3	35.3	Community	Male	Rural
C	1	30-Apr-2013	19	1	0.20	5.00	3	B	C	B	B	B	B	A	A	B	A	34	71.3	35.3	Community	Male	Rural
C	1	15-Oct-2013	24	1	0.20	5.00	3	B	C	B	B	B	B	A	A	C	A	39	53.2	24.9	Solidarity g	Male	Rural
D	1	24-Jun-2011	12	1	0.20	5.00	2	C	C	B	B	B	A	A	A	D	A	47	34.0	6.8	Solidarity g	Male	Urban
D	1	31-Dec-2013	29	1	0.20	5.00	2	E	C	B	B	B	B	A	B	B	B	63	4.8	0.7	Solidarity g	Male	Urban
E	1	14-Oct-2010	4	1	0.25	4.00	1	D	B	C	B	B	B	B	B	B	B	74	0.0	0.0	Individual	Male	Urban
E	1	1-Jan-2013	17	1	0.25	4.00	3	D	B	A	B	B	B	B	B	B	B	66	0.9	1.0	Individual	Male	Urban
F	1	17-Jun-2011	10	1	0.20	5.00	3	E	A	B	B	A	A	A	A	A	A	35	53.2	24.9	Community	Female	Rural
F	1	10-Jun-2013	20	1	0.20	5.00	3	E	A	B	B	A	A	A	A	A	A	35	53.2	24.9	Community	Female	Urban
G	1	3-Dec-2010	5	1	0.25	4.00	7	E	C	C	D	B	B	A	B	D	B	82	0.0	0.0	Solidarity g	Male	Rural
G	1	30-Oct-2013	25	1	0.25	4.00	7	C	C	C	D	B	B	A	B	B	B	60	4.8	5.2	Individual	Male	Rural
H	1	10-Nov-2011	14	1	0.20	5.00	5	D	A	C	D	B	B	A	B	C	B	64	4.8	0.7	Community	Male	Urban
H	1	11-Nov-2013	27	1	0.20	5.00	7	E	B	C	D	B	B	A	B	B	B	69	0.9	0.0	Community	Male	Urban
I	1	7-Dec-2011	15	1	0.20	5.00	2	E	B	B	D	B	B	A	B	D	B	75	0.7	0.0	Solidarity g	Female	Urban
I	1	9-Mar-2013	18	1	0.20	5.00	1	C	C	C	B	B	B	A	B	B	B	55	10.6	1.2	Solidarity g	Female	Urban
J	1	22-Dec-2010	6	1	0.25	4.00	2	D	C	C	C	A	B	A	B	C	B	64	4.8	0.7	Solidarity g	Female	Urban
J	1	5-Feb-2011	7	1	0.25	4.00	2	D	C	C	A	A	B	A	B	C	B	60	4.8	5.2	Solidarity g	Female	Urban
J	1	29-Aug-2011	13	1	0.25	4.00	3	D	C	C	D	B	B	A	B	C	B	69	0.9	0.0	Solidarity g	Female	Urban
J	1	17-Aug-2013	22	1	0.25	4.00	4	D	B	C	D	B	B	A	B	B	B	69	0.9	0.0	Solidarity g	Female	Urban
J	1	30-Oct-2013	26	1	0.25	4.00	4	E	B	C	D	B	B	A	B	B	B	77	0.7	0.0	Solidarity g	Female	Urban
K	1	30-May-2010	2	1	0.25	4.00	5	D	C	B	C	B	A	A	A	B	B	47	34.0	6.8	Community	Female	Rural
L	1	30-Jun-2012	16	1	0.15	6.67	7	C	B	B	B	A	B	A	B	A	B	39	53.2	24.9	Solidarity g	Female	Rural
L	1	29-Jul-2013	21	1	0.15	6.67	3	E	B	C	D	B	A	A	A	C	B	64	4.8	0.7	Community	Female	Urban
L	1	1-Jan-2014	30	1	0.15	6.67	4	D	B	C	D	B	B	A	A	C	B	60	4.8	0.7	Community	Female	Urban

For pro-poor organizations, the number of poor clients or people they serve (or the net number who leave poverty) matters more than than poverty rates (or changes in poverty rates).

As an extreme example, suppose that a small, subsidized pro-poor organization serves 100 clients, 100 percent of whom start poor and become non-poor by the time the organization runs out of subsidies and shuts down.

Suppose also that another larger, self-sustaining pro-poor organization has 10,000 new clients per year for 20 years. Ten percent of them are poor, and 10 percent of the poor become non-poor. Thus, 10,000 x 20 x 0.10 x 0.10 = 2,000 clients leave poverty.

The small, subsidized organization has a client-level poverty rate of 100 percent (versus 10 percent for the large, self-sustaining organization), but the small organization

serves fewer poor clients (100 versus 20,000) and has fewer poor clients leave poverty (100 versus 2,000). From the perspective of the well-being of people in the world and with all else constant, the second organization is preferred.

The point here is not that large, self-sustaining organizations are always better than small, subsidized organizations. Rather, the point is that a pro-poor mission may be better served with lower poverty rates, larger numbers of clients, or a longer organizational lifetime. This is why a pro-poor organization would want to look at not only poverty rates but also (and especially) numbers of poor clients.

More generally, decision-makers (whether managers, government, donors, or owners) can use scoring estimates to be more explicit and intentional in how they make trade-offs of depth of outreach (proxied by the poverty rate and its changes) against breadth of outreach (proxied by the number of poor clients and its changes).

Table 2 also shows:

- 80-percent confidence intervals for estimated poverty rates:
    — Between 23.7 and 28.3 percent for clients
    — Between 32.2 and 35.5 percent for people who live in households with clients

A confidence interval indicates the risk that bad luck-of-the-draw when sampling leads to an estimate that is far from the true value in the population. The narrower the interval, the lower the risk.

If new samples were to be repeatedly redrawn and re-scored, then 80 percent of the estimates of client's poverty rates would fall between 23.7 and 28.3 percent.

Confidence intervals for the number of poor clients (or people) in the population are found as the number of clients (or people) in the population, multiplied by the lower and upper limits of the corresponding poverty-rate confidence interval. In the example of Table 2, the 80-percent confidence interval for the estimated number of poor people in the population is 0.332 x 214 to 0.355 x 214, that is, 71 to 76.

All focal estimates should be reported with confidence intervals. This allows decision-makers to avoid placing too much weight on estimates that—due to small samples—have a high risk of being far from the true population value. Unfortunately, confidence intervals are not widely understood, and they are often not reported, especially for non-focal estimates. This is wrong, but perhaps not unwise, as more "table clutter" can detract from the impact of the analysis, giving users an excuse to tell themselves that it is all just statistical goobleygook that they do not need to heed.

Finally, the notes to Table 2 document:

- Sample sizes:
  — 10 clients
  — 33 people
- Population sizes:
  — 55 clients
  — 214 people

Table 3 is an example presentation of an estimate of a poverty rate at a point in time juxtaposed (via a cross-tab) with a non-scoring item (location as urban/rural).

Table 3: Example presentation of estimate of poverty rate at a point in time juxtaposed with a non-scoring item

Segment	Poverty rate (%)		Poor in pop. (#)		Dist. of sample (%)	
	Clients	Persons	Clients	Persons	Clients	Persons
Rural	57.8	57.6	14.3	77.6	35	52
Urban	8.5	7.8	2.6	6.2	65	48
Global:	26.0	33.9	14.3	72.4	100	100

Sample size (n) is 10 clients and 33 people.
Population size (N) is 55 clients and 214 people.

Like Table 2, Table 4 reports estimates of poverty rates for clients and people, now segmented by categories of a non-scoring item (urban and rural).[11] It also shows:

- Distribution of the sample:
  — Clients:
    ▪ 35 percent rural
    ▪ 65 percent urban
  — People who live in households with clients:
    ▪ 52 percent rural
    ▪ 48 percent urban

The global sample size is in the table's notes. The percentage distribution of the sample by segment shows each segment's relative importance better than would the number of clients or people in each segment.

---

[11] For context, the "Global" results from Table 2 also repeated in Table 3.

Table 5 is an example presentation of newly made-up results for estimates of changes in poverty through time based on two independent samples in two time windows.[12] Like Table 4, it juxtaposes the scoring estimates with location (urban/rural). In a real report, the global estimates of change through time—being focal estimates— would be emphasized and presented first without any cross-tabs (as in Table 2).

Table 5: Example presentation of estimates of the changes in poverty rates through time, juxtaposed with a non-scoring item, for two independent samples in two time windows

| Segment | Annual reduction in . . . | | | | Dist. of sample (%) at baseline | | Poverty rate (%) at baseline | | 80-percent confidence interval, annual reduction in pov. rate (% pts.) | | | |
| | Poverty rate (% pts.) | | Poor in population (#) | | | | | | Clients | | Persons | |
	Clients	Persons	Clients	Persons	Clients	Persons	Clients	Persons	Lower	Upper	Lower	Upper
Rural	2.0	1.6	0.4	1.8	35	52	57.8	57.6	1.8	2.2	1.3	1.9
Urban	1.0	1.5	0.2	1.5	65	48	8.5	7.8	0.9	1.1	1.3	1.7
Global:	1.7	1.6	0.6	3.3	100	100	26.0	33.9	1.3	2.0	1.4	1.8

Sample size (n) is 10 clients and 33 people at baseline. It is 20 clients and 97 people at follow-up.
Population size (N) is 55 clients and 214 people at baseline. It is 48 clients and 213 people at follow-up.
The baseline time window is calendar-year 2011. The follow-up time window is calendar-year 2014.

Considering only global estimates for the discussion here, Table 5 shows first:

- Annual reduction in poverty rates:
    — 1.7 percentage points for clients
    — 1.6 percentage points for people who live in households with clients

This means that the poverty rate for clients decreased from 26.0 percent in the baseline calendar-year of 2011 to 20.9 percent in the follow-up calendar-year of 2014. The reduction is expressed in units of "percentage points" (the difference in two percentages, here (26.0 – 20.9 percent) ÷ 3 years = 1.7 percentage points per year), not "percentages" (a ratio of a change divided by a baseline, here 1.7 ÷ 26.0 = 6.5 percent per year).

The adopted convention is to express change as a reduction, so positive numbers represent less poverty. Conversely, a negative reduction implies more poverty.

To be meaningful and comparable, changes are expressed in annual terms. For example, a 10-percentage-point reduction is huge in two years, smaller in two decades.

- Annual reduction in the number of poor in the population:
    — 0.6 clients
    — 3.3 people who live in households with clients

---

[12] Data in Table 3 is not used because this part of the analysis tool is still undeveloped. Therefore, figures in Table 5 may not be internally consistent.

The annual number of clients (or people) who leave poverty is estimated as the annual rate of reduction, multiplied by the number of population units at baseline. In the example of the rural segment in Table 5, the rate of reduction is 2.0 percentage points per year, and there are 55 clients, 35 percent of whom are rural at baseline. The annual number of rural clients who leave poverty is then 0.02 x 55 x 0.35 = 0.4.

Between the two time windows, clients can change segments (for example, moving from rural to urban), and households can gain or lose members. The analysis tool bases its estimates on the baseline values. Thus, Table 5 reports the baseline distribution of sampled clients and people by segment, as well as the baseline poverty rates. The baseline poverty rates also provide context for the estimated annual change in poverty rates.

Finally, estimates of change through time—and just like estimates at a point in time—have confidence intervals that should be reported.

Estimates of change through time, one time window, all clients scored more than once

The presentation of estimates of change through time with one time window and with all clients scored more than once is the same as shown in Table 5. While the technical details behind the derivation differ, the reporting format is the same.

This section showed how to report scoring estimates. The next step is analysis.

13. Do analysis, discussing possible cause-and-effect relationships that might inform decisions related with social performance

Beyond reporting, *analysis* is informed speculation about what scoring estimates might mean for business questions. Analysis compares estimates with benchmarks (to judge whether poverty outreach is high or low) and combines estimates with non-scoring resources (data, experience, theory, logic, and judgment) to help managers infer the causes of poverty outreach and then to seek potential ways to influence them.

This section discusses what analysis is and gives some examples.

Analysis is imperfect, but it needs to be done, and it can be done rigorously
Poverty scoring provides estimates that address questions central to the mission of pro-poor organizations:

- How many poor clients do we serve?
- How many poor clients are leaving poverty? At what rate?

Scoring does not reveal the causes behind the results that it measures. Yet the "why?" matters because the goal is not merely to track performance but also to help find ways to improve performance.

Decisions are based on beliefs about cause-and-effect, so informing decisions implies understanding causes. Poverty scoring *per se* does not reveal the causes behind its estimates, no more than a bathroom scale tells why weight decreased. But a manager—like a dieter—has other resources and can use them to help infer causality.

This is key: managers make decisions so as to improve performance. Decisions that are not expected to cause positive change are pointless. Even though poverty scoring by itself cannot show that an organization causes reductions in poverty, the organization uses poverty scoring mainly to help it find better ways to reduce poverty.

Managers are not social scientists. Managers have a narrow goal: make a good-enough decision to improve organizational performance. Academics have a broader goal: discover general truths. Managers must "do something" on a deadline to address a context-specific problem with incomplete information and a limited budget, while academics have more time to seek an ideal more closely. Both views have their place, but they are not the same.

In their lives, people are both managers and academics, but more often managers. They make some big-stakes decisions (say, whether to drink milk at all) that justify costly, lengthy decision-making. But most of the time, the stakes are lower (say, whether to buy more milk now or to wait a few days). Rather than wastefully insist on perfect decisions, people use judgment to balance the benefit of a better decision against

the cost of better decision-making, accepting that some decisions will—in hindsight—not turn out to have caused the hoped-for impact on their lives.

Individual and organizational decision-making is "good-enough-for-government-work" because it makes sense. It is normal to act on uncertain beliefs about cause-and-effect. What matters is that the analysis process make reasonable attempts to reduce uncertainty, use facts, and make the process transparent and thus open to improvement. A decision is *rigorous* not because it is derived from quantitative data and is almost certainly correct but rather because it makes its judgments and assumptions transparent, discusses the uncertainty of the result (and possible consequences of misjudgment), and balances the benefits and costs of the whole process.

On its own, poverty scoring's estimate of a change does not establish cause-and-effect with certainty. But scoring cannot inform decisions unless managers believe that an estimated change is, to some degree, an effect caused by a pro-poor organization. To budge from the status quo requires some leap from correlation to causation, and that is fine as long as the process is careful and the uncertain issues are made explicit.

For example, a poverty-scoring analysis might state: "We speculate that about half of the estimated reduction in poverty is due to our organization because the reduction is twice that of our region, because we aim to reduce poverty, and because 90 percent of clients in a qualitative survey said that they are 'a lot' better off with us than they would be without us."

This judgment is rigorous because its assumptions and leaps-of-logic are explicit. It may be off, but it is open-source and so can be improved. *Rigorous* doesn't mean incontrovertible; it means controvertible, as the "iffy" parts are discussed. Being careful and transparent when making decisions and informing beliefs is the essence of both good social science and good management.

Poverty scoring aids in this process by inexpensively quantifying clients' poverty (and its uncertainty). More broadly, scoring fosters an organizational culture of transparent, intentional decision-making, especially in terms of social performance.

Like all "decision tools", poverty scoring does not make decisions. Analysis that compares scoring estimates with benchmarks and combines estimates with non-scoring resources will—with luck—point toward likely cause-and-effect relationships that organizations can take advantage of to improve social performance.

Comparing scoring estimates with benchmarks

Suppose an organization's estimated poverty rate for new clients is 25 percent; is this high or low, impressive or disappointing? Meaningful analysis must establish a standard or benchmark for comparison. Without a benchmark, a poverty rate of x percent can be deemed adequate just as reasonably as it can be deemed inadequate.

But what benchmark to use? Common, reasonable ones are:

- Past estimates (to check whether there is improvement with time)
- Goals set by an organization (to check whether targets are met)
- An organization's mission statement, internal beliefs, or external claims (to check whether they are in accord with reality)
- Regional or national estimates (to check whether client households are more or less likely to be poor—or to exit poverty—than the average household)

Setting benchmarks can prompt pro-poor organizations to think intentionally about their mission and what exactly it means "to serve the poor". Is it that the number of poor clients exceeds zero? That the poverty rate of new clients exceeds the population average? That 90 percent of new clients are poor? That clients' poverty rate exceeds that of clients of non-pro-poor organizations that provide similar services?

As an example of benchmarked analysis, consider VisionFund/Ecuador.[13] VFE's main strategic partner explicitly expects new clients to have a poverty rate at least as high as that of Ecuador as a whole.[14] The all-Ecuador poverty rate at the end of 2013 at the household level was about was 20.1 percent,[15] while VFE's new-client poverty rate was about 29.0 percent (Páez, 2014). Thus, VFE beat its benchmark by about 9 percentage points.[16]

VFE has not set an explicit goals for rates of poverty reduction, but a defensible benchmark is that the reduction for its clients be at least as great as the reduction for households the country as a whole. From 2011 to 2013, the all-Ecuador poverty rate at the household level fell by about 1.9 percentage points per year (INEC, 2014), while the annual reduction for VFE's clients—based on a sample of 402 clients in the time window from 2011 to 2013 in which each sampled client was scored twice—was about 3.3 percentage points per year (Table 6).

---

[13] `visionfund.org/2109/where-we-work/latin-america/ecuador/`, retrieved 24 October 2014.

[14] Being *poor* is defined as having consumption is below the national poverty line.

[15] To make INEC's (2014) person rate comparable with VFE's household rate, the relative difference between the all-Ecuador rates for people and households in the 2005/6 *Encuesta de Condiciones de Vida* (ECV) is applied to the 2013 person rate.

[16] The comparison is imperfect because the scorecard measures consumption-based poverty based on the 2005/6 ECV but INEC (2014) reports income-based poverty in the 2013 *Encuesta Nacional de Empleo, Desempleo y Subempleo.*

Table 6: Estimates of annual rates of reduction in poverty rates through time, VisionFund/Ecuador

Segment	Annual reduction in . . .			Poverty rate (%) at baseline	80-percent confidence interval, annual reduction in pov. rate (% pts.)	
	Poverty rate (% pts.)	Poor in population (#)			Clients	
	Clients	Clients	Persons	Clients	Lower	Upper
Global:	3.3	2,120	8,692	32.8	3.2	3.3

Sample size (n) is 402 clients and (assumed) 2,010 people.
Popualtion (N) is assumed to be 65,000 clients and 260,000 people.
The time window is 2011 to 2013.
Only estimates derived with VFE displayed.

Of course, this does not necessarily imply that the 3.3 – 1.9 = 1.4 percentage-point annual difference was caused by VFE. In the absence of a control group that resembles VFE clients in all ways except the use of VFE (or in the absence of strong assumptions), poverty scoring cannot tell why poverty changed, just as a bathroom scale cannot tell whether a person lost weight due to a regimen of diet and exercise or whether he/she just took off his/her shoes. VFE's impact could be larger or smaller than 1.4 percentage points per year (or zero, or negative). Nevertheless, the fact that VFE's clients improved faster than Ecuador as a whole encourages the hope that VFE had a positive impact.

A rigorous estimate of VFE's impact on poverty can be had by explicitly assuming that its clients are a random sample from Ecuador's population. Given this, the full client/population difference of 1.4 percentage points per year can be attributed to being a client of VFE. Of course, being rigorous does not mean incontrovertible; it is unlikely that VFE caused the full 1.4-percentage-point difference (it is also unlikely that VFE had a no or negative impact). Instead, *rigorous* means that discussion meant to improve the estimate can focus on the assumption that VFE clients are representative of Ecuador's population as a whole.

Juxtaposing scoring estimates with non-scoring items and other resources to prospect for possible causes of poverty outreach

The analysis process does not end with the development of beliefs about causes. Managers also need to find ways to influence causes and then judge whether it is worthwhile to try to do so.

The main method is to cross scoring estimates with non-scoring items and then—considering experience, theory, common sense, and other dimensions of knowledge—judge the extent to which observed correlations represent causes. As usual, scoring estimates do not announce their own import for social-performance questions. Nevertheless, their correlations may reinforce (or weaken) existing beliefs or point to potentially high-return areas for further investigation. The analyst's job is to stake out the most promising possible veins of ore for managers to mine.

The first analyses should cross scoring estimates with non-scoring items that are under the organization's direct control such as:

* Type of services used by a client
* Characteristics of a given type of service used by a client
* Field office and field agent associated with a client

For example, VFE crossed the type of loan with the estimated annual rate of reduction in poverty, finding that borrowers with community banks exit the fastest, that most borrowers who exit poverty are members of solidarity groups, and that individual borrowers exit the slowest (Table 7).

Table 7: Estimates of annual rates of reduction in poverty rates through time of VFE clients by type of loan

	Annual reduction in . . .	
	Poverty rate (% pts.)	Poor in population (#)
Type of loan	Clients	Clients
Community bank	4.1	904
Solidarity group	3.0	1,204
Individual	0.6	12
Global:	3.3	2,120

Sample size (n) is 402 clients.

Population (N) is assumed to be 65,000 clients.

The time window is 2011 to 2013.

Only estimates derived with VFE displayed.

Analysis of Table 7 suggests some potentially useful questions and possibilities:

* Clients of community banks and solidarity groups left poverty the fastest, consistent with VFE's strategy to focus on these two types of loans
* Clients of community banks have the fastest rate of reduction, but more solidarity-group clients left poverty, sparking follow-up questions:
    — Should VFE try to expand community banks, solidarity groups, or both?
        ▪ Why do solidarity groups have more clients than community banks?
        ▪ How could VFE expand either loan type, if it wanted to?
        ▪ What adjustments would attract those who did not join before?
        ▪ What matters more, the rate of exit or the number of clients who exit?

— Most community-bank clients are rural women in farming households. Is the faster reduction in poverty due to the type of loan or the type of client?
- If the client, how can VFE reach/attract more rural women?
- If the loan:
  — How do loan types differ?
  — How do these differences cause different rates of exit?
  — Can the causal features be grafted from loan type to another?
— Do individual borrowers contribute to VFE's pro-poor mission?
- Do they employ poor people?
- Do they contribute disproportionately to VFE's sustainability?

As in this example, analysis often merely points managers to additional questions, which, if addressed, might provide concrete ways to attract more poor clients—and poorer clients—and then to help them to exit poverty faster.

As another example, differences in poverty by region may be driven by causes beyond the influence of the organization (other than the choice of where put field offices). If, however, they are not due to immutable regional characteristics (such as climate or history), then they may be due—for example—to differences in the ability/effort of the organization's personnel. Perhaps managers could seek to find out what high-performing regions do differently and then attempt to transfer these things to lower-performing regions.

Pro-poor organizations would also like to know how poverty is linked with "failure", whether that means dropping out soon after joining, repaying late, or taking a class but not changing behavior. Such knowledge could guide efforts to improve value to clients and reduce "failure". Is Nobel prize-winner Muhammad Yunus of the Grameen Bank in Bangladesh correct when he says that "the poor always pay back" (Dowla and Barua, 2006)? If not, why not? How well the poor repay may depend on how well a loan fits their situation. If drop-out is higher among the poor, is it because they are poor, or is it because aspects of the loans that they get are not well-tailored to the poor?

In an example analysis that juxtaposes poverty, age, sex, and marital status, Mokoena, Páez, and Schreiner (2014) find that the annual rate of reduction of poverty among VFE clients ages 20 to 29 is slower (about 1 percentage point, Table 7) for never-married women but faster (about 6 percentage points) for married women. In contrast, the estimate for young, never-married men is close to the all-client average (about 3 percentage points).

Table 8: Estimates of annual rates of reduction in poverty rates through time for women and men clients of VFE ages 20 to 29 by marital status (married and never-married)

Segment	Annual reduction in poverty (percentage points)	Share of sample (%)
Women		
Married	6.2	9
Never-married	1.1	15
Men		
Married	3.5	1
Never-married	2.9	11

Note: Clients in this table were 20 to 29 years old when they joined VFE in 2010 and 2011. They were scored at in-take and then again in 2013. They are part of a random sample of 402 clients from the population of new clients in 2010 and 2011 who were still active in 2013. The table omits clients who were 30 or older or whose marital status was something other than married or never-married.

Of course, VFE cannot induce young, female clients get married, nor can it make them age faster. It might, however, be able to learn how its services affect poverty by asking about the causes behind the correlations in Table 7. In particular, perhaps many married and never-married women have children, but married women leave poverty faster because they can spend more time in their (VFE-supported) businesses because their husbands help. If the job of raising children born out of wedlock falls mostly on the mother, then child-care responsabilities and the lack of a second income could plausibly cause a slower rate of exit from poverty for young, never-married women.

This matters to VFE because its mission focuses on women and children and because young, never-married women are one-seventh of its clients. Of course, scoring analysis alone cannot tell VFE what to do, and managers must balance many considerations. Should VFE discourage young, never-married mothers from borrowing? Would a new product—such as a matched savings account (Schreiner, 2004)—be better than a loan for this segment? Would it help if VFE hired lawyers to support the legal process of marrying children's fathers or enforcing child-support obligations? While poverty scoring cannot answer these questions for VFE, it does point to an area of inquiry where new insights might directly improve its fulfillment of its pro-poor mission.

14. Present analysis clearly, highlight focal estimates, and discuss caveats

All the work so far counts for naught if it is not presented well. The strength of poverty scoring is its straightforwardness and transparency; if poverty-scoring analysis is to affect decisions, then it too must be presented straightforwardly and transparently.

The first step is simply to highlight the focal estimates—without analysis—on the first page or first paragraph. Some scoring reports "bury the lead" or never even report estimates of (rates of change in) poverty rates or of the number of clients (or people) who are poor when they start and who leave poverty during their tenure. The main impact of poverty scoring on social performance usually comes simply from seeing a quantitative measure of poverty outreach. Make it so managers cannot miss it.

Second, follow the example table formats here. They present the focal estimates first; they have simple, clear, accurate labels; and they include other estimates that put the focal estimates in context. Do not get too fancy nor re-invent the wheel.

Third, do not expect managers to figure out what scoring estimates mean. Instead, tell them what you (the analyst) speculate that the estimates might imply. Do not worry about being wrong or overstepping your bounds; managers cannot study the tables as carefully as the analyst, they are used to delegating, and they welcome advice about what it all might mean. They can respond and build on to your tentative inferences more readily than they can come up with their own. And if they have to work to digest the report, then they may just skim it and go on to the next thing. Take the analysis as far as you can, to the point where they can combine it with their expertise, experience, judgment, and knowledge of the organization's context and strategy to guess at changes that might improve social performance.

Fourth, make caveats clear, but do not succomb to the desire for certainty. Keep in mind that while some speculations about cause-and-effect do not hold true, others do. Scoring estimates cannot be taken at face value, but they do have some value. Correlation does not equal causality, but correlation usually signals some causality, and analysts and managers should pool their complementary knowledge bases to make their best guess of how far they can go. Managers—and everyone else—do this all the time without a second thought. When looking at "hard numbers", however, people can forget that wise decision-making—like science—is not quantitativeness and certainty to the exclusion of human judgment but rather careful process and openness to improvement. The strength of poverty scoring is not the perfection of its estimates. Instead, it is that its low cost allows its imperfect estimates to have a chance at being worthwhile in decision contexts where time and other resources are scarce.

Fifth, document the details of implementation in an annex. Recording how the work was done helps to establish (for example) the representativeness of the sample and (in general) the extent to which the data can be trusted.

Finally, use exemplary scoring-analysis reports as guides.

References

Dowla, Asif; and Barua, Dipal. (2006) *The Poor Always Pay Back: The Grameen II Story*, Bloomfield, CT: Kumarian Press.

INEC. (2014) "Indicadores de Pobreza", Quito, `ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/Empleo-dic-2013/Presentacion_Pobreza_Dic_2013.pdf`, retrieved 16 August 2014.

IRIS Center. (2007) "Manual for the Implementation of USAID Poverty Assessment Tools", `pdf.usaid.gov/pdf_docs/PNADQ620.pdf`, retrieved 5 September 2014.

Mokoena, Refilwe; Páez, Wilman; and Mark Schreiner. (2014) "Resumen de la Semana de Consultoría: Capacitación en el Uso del PPI para el Seguimiento de Cambios y Desarrollo de Herramientas y Procesos para el Uso en la Red de Visión Mundial", 8 August, presented to the upper management of FODEMI, Ibarra.

Páez, Wilman. (2014) "Informe PPI, periodo comprendido entre enero y junio 2014", FODEMI, Ibarra.

Schreiner, Mark. (2004) "Asset-Building for Microenterprise through Matched Savings", `microfinance.com/English/Papers/IDAs_for_Microenterprise.pdf`, retrieved 24 May 2012.

_____. (2002) "Aspects of Outreach: A Framework for the Discussion of the Social Benefits of Microfinance", *Journal of International Development*, Vol. 14, pp. 591–603.

Toohig, Jeff. (2008) "Progress Out of Poverty Index: Training Guide", Grameen Foundation, `microfinancegateway.org/gm/document-1.1.6364/PPITrainingGuide.pdf`, retrieved 17 June 2013.