Simple Poverty Scorecard® Tool
Tanzania

Mark Schreiner

20 August 2021

Waraka huu na zana ya kukusanya data ziko kwa lugha ya Kiswahili kwenye scorocs.com
This document and a data-collection app are in English at scorocs.com

The Scorocs Simple Poverty Scorecard-brand poverty-assessment tool is a low-cost, transparent way for pro-poor programs in Tanzania to get to know their participants better so as to prove and improve their social performance. Responses to the scorecard's 10 questions can be used to:

- Check poverty rates and numbers of poor people among in-coming participants
- Track changes in poverty among on-going participants
- Segment participants for differentiated treatment based on poverty

Version note

The new scorecard for Tanzania is based on data from 2018. It replaces the old scorecards in Schreiner (2016a and 2013a) that were based on data from 2011/12 and 2007. The new scorecard should be used from now on. Estimates are not comparable across the three scorecards because they use different definitions of *poverty*. Thus, estimates of changes in poverty over time must use the same scorecard for both baseline and follow-up.

Acknowledgements

Chalice funded this paper, and Tanzania's National Bureau of Statistics provided the data. Thanks go to Novati Buberwa, Albina Chuwa, Nadia Belhaj Hassine Belghith, Alisha Lebouthillier, MaryAnne McKinnon, Sylvia Meku, Rose Mungai, and Richard Valliant.

Scorocs® Simple Poverty Scorecard® Tool

			Name	Identifier
Interview ID:				
Interview date:		Direct participant:		
Country:	TZA	Field agent:		
Scorecard:	003	Service point:		
Sampling weight:		Number of household members:		

Question	Response	Points
1. In which region does the household live? (*Based on the enumerator's knowledge*)	A. Ruvuma, or Lindi	0
	B. Dodoma, Mwanza, Rukwa, or Shinyanga	4
	C. Tanga, Morogoro, Arusha, or Iringa	6
	D. Pwani, Geita, Simiyu, Manyara, Njombe, or Katavi	7
	E. Kigoma, Mtwara, Singida, Tabora, or Songwe	8
	F. Dar es Salaam, Kagera, Mbeya, Kilimanjaro, or Mara	12
2. How many members does the household have? (*Based on the "Back-page Worksheet"*)	A. Eight or more	0
	B. Seven	5
	C. Six	6
	D. Five	9
	E. Four	13
	F. Three	19
	G. Two	28
	H. One	40
3. What is the main building material of the floor?	A. Earth/sand, dung, vinyl or asphalt strips, wood planks, palm/bamboo, or other	0
	B. Cement, ceramic tiles, or parquet or polished wood	6
4. Does the household own any charcoal stoves?	A. No	0
	B. Yes	3
5. Does the household own any tables?	A. No	0
	B. Yes	3
6. Does any member of the household own a television?	A. No	0
	B. Yes	12
7. Does any member of the household own a mobile phone?	A. No	0
	B. Yes	6
8. Does any member of the household own any cattle, goats, sheep, or hogs?	A. No	0
	B. Yes	3
9. In the past week, did the household consume any chicken, goat, fish, beef, or other meat?	A. No	0
	B. Yes	7
10. How many meals does the household usually have per day?	A. Two or less	0
	B. Three or more	8

Score:

Back-page Worksheet: Members of the Household

Fill out the scorecard header first. Include the interview's unique identifier (if known), the interview date, and the sampling weight of the participating household (if known). Then record the full name and the unique identification number of the direct participant (who may differ from the respondent), of the direct participant's field agent (who may differ from you the enumerator), and of the service point that the direct participant uses (if any and if known). Circle the response to the first scorecard question ("In which region does the household live?") without asking the respondent.

Then read to the respondent: *Please tell me the first name (or nickname) of each household member. A* household *is one person alone or a group of people (regardless of blood or marital relationships) who share consumption, pool their resources together, and have lived in the same residence for at least two weeks*.

Write down the name of each member, beginning with the head and his/her spouse (if there is one). Record the number of household members in the scorecard header next to "Number of household members:". Then circle the response to the second scorecard question about the number of household members.

Read the remaining eight questions aloud. Always keep in mind and apply the detailed instructions in the "Interview Guide".

First name or nickname?
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.
13.
14.
15.
16.
17.
Number of household members:

Figure 1: Conversion of scores to poverty likelihoods

	Poverty likelihood (%)														
	National				Intl. 2011 PPP				Percentile-based lines						
Score	Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
0–17	40.0	71.4	94.8	98.4	95.0	98.9	100.0	100.0	46.3	62.7	88.9	95.0	97.5	98.9	100.0
18–21	19.8	60.1	89.8	96.3	85.4	98.0	100.0	100.0	26.2	50.0	75.3	85.4	93.4	98.0	100.0
22–24	19.4	58.2	86.5	95.9	84.2	96.5	99.5	100.0	25.8	48.5	74.4	84.5	91.4	97.0	99.5
25–26	17.5	53.8	81.9	92.5	79.8	96.5	99.5	100.0	18.5	39.9	71.0	79.9	86.4	97.0	99.5
27–28	12.2	46.1	81.1	92.1	79.7	96.3	99.4	100.0	15.4	34.7	67.9	79.7	85.3	97.0	99.2
29–30	12.2	36.9	78.1	91.6	73.3	95.9	99.4	100.0	15.4	30.2	61.1	75.2	84.4	96.4	98.9
31–32	9.2	32.3	73.3	91.6	69.3	95.9	99.3	100.0	12.3	25.3	59.3	70.6	82.3	96.4	98.4
33–34	7.0	31.8	72.8	89.6	65.7	94.0	99.2	100.0	9.8	24.1	52.5	68.9	80.7	95.3	98.4
35–36	6.0	28.8	68.0	87.1	60.3	91.6	97.3	100.0	8.0	19.8	45.6	62.4	75.5	93.2	97.1
37–38	6.0	25.4	63.0	85.2	54.4	90.4	97.3	100.0	5.8	16.9	41.2	56.3	72.8	92.5	96.9
39–40	3.9	19.4	58.1	81.0	48.2	87.8	97.3	100.0	4.1	9.8	32.1	49.4	67.3	90.9	96.5
41–42	3.9	19.4	56.1	78.1	45.8	86.4	97.3	100.0	4.1	9.8	32.1	47.3	63.6	88.9	96.5
43–45	1.7	17.6	53.0	75.4	42.6	85.2	97.1	100.0	3.0	8.9	29.7	43.9	60.0	87.2	93.5
46–48	1.1	12.1	43.1	70.2	33.6	80.5	95.2	100.0	1.6	6.4	21.3	36.4	53.5	83.0	91.9
49–51	1.1	8.1	32.7	60.8	24.7	70.4	92.1	99.8	1.3	4.1	15.7	26.3	41.4	75.1	87.8
52–54	0.9	5.9	25.5	52.8	17.6	57.7	92.1	99.7	1.1	3.2	10.1	18.2	30.0	64.5	85.8
55–57	0.4	4.6	22.1	48.5	14.9	52.9	90.5	99.7	0.2	1.9	8.5	15.6	29.5	60.4	85.1
58–61	0.1	1.8	14.0	36.7	7.8	42.2	81.2	99.7	0.2	1.0	5.7	8.3	17.0	51.6	75.4
62–66	0.1	0.7	8.9	26.9	5.3	35.9	74.9	99.6	0.0	0.4	3.6	5.7	11.5	43.4	68.2
67–100	0.1	0.3	2.5	9.8	1.5	13.7	43.2	95.3	0.0	0.3	0.8	1.5	3.4	16.3	35.2

Figure 2: Errors in estimated snapshot head-count poverty rates in a single time period, along with margins of error and the α factor for finding margins of error and sample sizes

	Poverty lines														
	National				Intl. 2011 PPP				Percentile-based lines						
	Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
Estimation error	–0.9	–1.5	–1.8	+2.1	–2.8	+4.0	–0.9	–0.1	–1.0	–1.4	–5.2	–2.6	–0.2	+1.3	–2.2
Margin of error	1.8	2.8	2.9	2.7	2.8	2.6	1.3	0.1	1.9	2.5	3.1	2.8	2.8	2.4	1.4
α factor	1.52	1.37	1.14	1.11	1.14	1.12	0.80	0.27	1.41	1.41	1.26	1.13	1.09	1.07	0.81

Estimation errors from the scorecard with 1,000 bootstrap samples of n = 16,384 households from the validation sample.
Estimation errors are average differences between estimates and observed values, in percentage points.
Margins of error are ± percentage points with 90-percent confidence for samples of n = 1,024.
α is an average across 1,000 bootstrap samples of n = 256, 512, 1,024, 2,048, 4,096, 8,192, and 16,384.

Table of Contents

Figures

Scorocs® Simple Poverty Scorecard® Tool
Tanzania

1. Introduction

The Scorocs Simple Poverty Scorecard-brand poverty-assessment tool for Tanzania is a low-cost, transparent way for pro-poor programs to get know their participants better so as to prove and improve their social performance.

1.1 Questions addressed by the scorecard

To address the question of "How many poor people does our program attract?", the scorecard can take a snapshot in a single time period with a census or a sample of in-coming households to estimate both head-count poverty rates and the number of poor people.

To address the question of "How has poverty changed for on-going participants?", the scorecard can be applied across two time periods with samples from a given population of on-going participants to estimate both net annual changes in head-count poverty rates and net annual changes in the number of poor people.

The scorecard can also be used for targeting, that is, to segment participants for differentiated treatment based on poverty.

It is difficult and costly for pro-poor programs to address these questions with the traditional direct approach to poverty assessment via consumption surveys. A case in point is the 2018 Household Budget Survey (HBS) by Tanzania's National Bureau of Statistics (NBS). The 2018 HBS has 62 pages and asks more than 600 top-level questions, many of which have several follow-up questions or are repeated (for example, for each household member or for each crop).[1]

[1] Enumerators also made 14 visits to each interviewed household to help compile a consumption diary.

1.2 How the scorecard works

The scorecard has 10 factual questions that are drawn from the exhaustive 2018 HBS. Examples include: "What is the main building material of the floor?" and "Does the household own any charcoal stoves?".

The 10 questions are selected to be:

- Inexpensive to collect, easy to answer quickly, and straightforward to verify
- Strongly and intuitively linked with poverty
- Liable to change over time as poverty changes
- Applicable in all regions of Tanzania[2]

Each question has multiple-choice response options, with points assigned to each response. The points are zeroes or positive whole numbers. The points are derived from the statistical links between responses and consumption-based poverty in the 2018 HBS.

Adding up the points for a given household gives a *score* that ranges from 0 to 100. The lower the score, the poorer the household.

An enumerator can interview a household, record its responses on paper or on a hand-held device, and add up the household's score (if needed for on-the-spot segmentation) in about ten minutes.[3]

Back at the office or in the cloud, a household's score is converted into an estimated probability (the *poverty likelihood*) that the household is poor for a given poverty line. The links between scores and poverty likelihoods are based on HBS data.

The average of poverty likelihoods across the members of sampled households is an estimate of the head-count poverty rate among people in the sampled population.

This estimated poverty rate may then be used to estimate:

- The number of poor people in in-coming households in a single time period
- The net number of poor people in households of on-going participants who rise above a poverty line across two time periods

[2] Except Zanzibar, which is not included in the 2018 HBS.
[3] Responses on paper are entered in a spreadsheet or database later at an office.

1.3 Targeting

The scorecard can also be used to segment participating households for differentiated services. Unlike some other targeting tools—such as the World Bank's "proxy-means tests"[4]—the scorecard is transparent, freely available,[5] and tailored to the capabilities and purposes not of national governments but rather of local pro-poor programs. The feasible poverty-assessment tools for such programs are typically blunt (such as rules based on land ownership or housing quality) or subjective and relative (such as community-based, participatory wealth ranking facilitated by skilled field workers). Poverty assessments based on these approaches may be costly, their accuracy is unknown, and they are not comparable across places, programs, nor time.

1.4 Consumption-based poverty

Tanzania's scorecard is a quantitative way to assess whether a program's participants have consumption expenditure below any of 15 poverty lines, for example:

- Tanzania's official national (basic-needs) line of TZS1,619 per adult equivalent per day, giving a country-wide head-count poverty rate of 26.4 percent in 2018
- The World Bank's "lower-income poverty line" of $1.90 per person per day 2011 PPP (TZS1,794), giving a poverty rate of 48.9 percent

A program uses only the poverty line(s) that fit its context and mission. For example, a program may report poverty estimates to funders based on a World-Bank international line while internally using a national line or percentile-based line.

1.5 Transparency

The scorecard's design aims to make its workings clear to program managers. The tool's adoption stems from the low cost of its short interviews and from the fact that managers can see for themselves how the scorecard works and that its approach makes sense. Similar tools have been around for decades, but pro-poor programs have rarely used them. This is not because these tools are inaccurate, but because *how* they work is unclear or hidden.

---

[4] Coady, Grosh, and Hoddinott, 2004.

[5] Tanzania's scorecard is not in the public domain; it is copyright © 2020 Scorocs.

When scorecard projects fail, the cause is not usually inaccuracy but rather a program's failure to commit to the work-a-day project management needed to integrate the scorecard in the program's processes and to train and convince employees to use the tool properly.[6] For tool-based estimates of social outcomes such as poverty, data scientists have long known that there is almost no trade-off between the straightforward and transparent versus the complex and opaque.[7] Project risk is less technical and more human, not statistics but organizational-change management.

1.6 Assumptions and estimation errors

Like all predictive tools, the scorecard makes two fundamental assumptions:

- The scored sample is representative of the same population as that whose data was used to construct the scorecard
- The links between responses and poverty are the same in the scored sample as in the population whose data was used to construct the scorecard

Of course, the assumptions do not hold to some unknown degree.[8] In particular:

- A given program's participants are not representative of Tanzania overall
- Over time, the links between responses and poverty drift or shift

Scorecard estimates have errors because the scorecard incorrectly acts as if the links between responses and poverty in all scored samples and in all time periods are the same as in the construction data from the 2018 HBS. Reality diverges further from assumptions as:

- More time passes since the collection of construction data
- A program's participants differ from the country's general population
- Attrition has changed the composition of a cohort of on-going participants
- Change has been rapid (say, due to war, plague, or changes in the program itself)[9]

---

[6] Schreiner, 2002.

[7] Dupriez, 2018; Caire and Schreiner, 2012; Schreiner, 2012; Hand, 2006; Lovie and Lovie, 1986; Kolesar and Showers, 1985; Stillwell, Barron, and Edwards, 1983; Dawes, 1979; Wainer, 1976; Myers and Forgy, 1963.

[8] Diamond *et al.*, 2016; Tarozzi and Deaton, 2009.

[9] For example, the 2020/21 economic downturn due to COVID–19 changed the links between poverty and questions, but the Tanzania scorecard still uses 2018 links.

For any particular scorecard and scored sample, the estimation error due to migration away from the assumptions is unknown. It is known, however, that the scorecard's targeting is robust. That is, the extent to which assumptions diverge from reality is not strongly linked with the extent to which the scorecard gives lower scores to more-poor households and higher scores to less-poor households. It is also known that the scorecard's estimation errors are larger when estimating changes in poverty across two periods (or across two scorecards) than when estimating poverty in one period.

There are no rules nor formulas that automatically signal when estimation error is too large for estimates to be useful. Program managers must make their own judgments based on common sense and on what they know about their context and their participants from non-scorecard sources.

In practice, scorecard estimates often serve as a basic check on whether a pro-poor program is indeed *pro-poor*. The estimates address existential questions such as:

- "How many in-coming participants are below the national poverty line?"
- "Are in-coming participants poorer than the average person in the area where we work?"
- "Are our poor participants more likely to rise above a poverty line than the average poor person in the area where we work?"

For such existential checks on whether a program lives out its purported social mission, estimation errors will often be small enough to be immaterial.

1.7 Estimation errors when assumptions hold

If the scorecard's assumptions do hold, then the scorecard estimators are statistically *unbiased*. That is, the true value in the population matches the average of scorecard estimates from repeated samples.

The assumptions do hold when the scorecard is tested against households in the validation sample from the 2018 HBS that is not used to construct the scorecard. Smaller errors in this ideal case imply smaller-than-otherwise errors in real-world use.

Even so, there are estimation errors on average in the validation sample because there is only one scorecard, and it is derived from one construction sample and applied to a single validation sample. Figure 2 documents the error for snapshot estimates of poverty rates in one time period, allowing scorecard users to adjust for this error.

1.8 What is next?

Section 2: How to convert responses into poverty likelihoods

Section 3: How to calculate scorecard estimates

- Snapshot estimates of:
  - Head-count poverty rates in a single time period
  - Number of poor people in a single time period
- Estimates across two time periods in:
  - Annual net change in poverty rates with one sample scored twice
  - Annual net change in the number of poor people with one sample scored twice
  - Annual net change in poverty rates with two independent samples
  - Annual net change in the number of poor people with two independent samples

Section 4: How to design scorecard surveys and samples

Section 5: How to use scores for targeting

After Section 5, the "Interview Guide" tells how to ask questions—and how to interpret responses—so as to mimic practice in Tanzania's 2018 HBS as closely as possible. The "Guide" (and the "Back-page Worksheet") are integral parts of the scorecard. Do not ignore them.

The annexes provide details for advanced users:

Annex 1: Data used for construction and validation

Annex 2: Definition of poverty

Annex 3: Scorecard construction

Annex 4: Estimates of poverty likelihoods

Annex 5: Error and margins of error

Annex 6: Formulas for sample size

Details on cited References appear at the end.

2. How to convert responses to poverty likelihoods

This section tells how to:

- Collect a household's responses to scorecard questions
- Convert responses to points
- Add up points to get scores
- Convert scores to poverty likelihoods

The next section tells how to combine poverty likelihoods from a sample of households to estimate poverty.

2.1 Instructions for enumerators

An *enumerator* asks a scorecard's questions to a respondent and then records the responses. An enumerator may or may not be same as the program's field agent (if any) who is associated with a participating household.

Enumerators should interview a sampled household at the household's residence using an app on a hand-held device or a paper scorecard along with the "Back-page Worksheet". Following the "Interview Guide", enumerators should:

- Record administrative information in the scorecard header:
  - Interview identifier (if known)
  - Interview date (required)
  - Country code ("TZA", pre-filled)
  - Scorecard code ("003", pre-filled)
  - Sampling weight assigned to the household by the survey design (if any and if known)
- Record names and identifiers (if known) in the scorecard header:
  - *Direct participant*. This is the adult household member who directly interacts with the pro-poor program. He/she may or may not be the same as the respondent who responds to the scorecard questions. For example, a direct participant with a microfinance program is a borrower or a saver, and a direct participant with a child-health program is a child's parent or guardian
  - *Field agent* (if there is one). This is the direct participant's main, repeated point of contact with the program. The field agent may or may not be the same as the enumerator. For example, the field agent in a microfinance program is a loan officer or savings collector, and the field agent in a child-health program is a community health-care worker

  - *Service point* (if there is one). This is the program office that is relevant to the direct participant. The service point is usually the base of operations of the direct participant's field agent (if there is one) or where the direct participant usually goes to do program business. For example, the service point for a microfinance program is a branch, and the service point for a child-health program is a health post
- Mark the response to the first scorecard question ("In what region does the household live?"). If the enumerator already knows the region (as is usually the case), then the question need not be asked directly of the respondent
- Complete the "Back-page Worksheet" with each household member's first name (or nickname)
- If using a paper scorecard, then use the "Back-page Worksheet" to record:
  - The number of household members in the header next to "Number of household members:"
  - The response to the second scorecard question ("How many members does the household have?")
- Read the remaining eight questions aloud one-by-one and in order, marking the responses given by the respondent
- When marking a response on paper, write each point value in the far right-hand column. Then make single circle around the pre-printed response, the pre-printed points, and the hand-written points. This helps to reduce later data-entry mistakes
- Add up the points to get the score (if needed on-the-spot and if using a paper scorecard)
- Implement targeting policy (if any) based on the score
- Upload the data with a mobile data-collection tool, or deliver the filled-out paper scorecard to a central office for data entry, reporting, and analysis

2.2 Header, 'Back-page Worksheet', 'Interview Guide', and audits

Fill out the scorecard header as best you can; do not skip it. Scorecard estimates are more useful if they can be linked—via names or identifiers—to a program's existing data on direct participants, field agents, or service points. Record the types of identifiers that are used in the program's databases, be they program-specific or government-issued. Be sure to record the number of household members not only indirectly via the scorecard's second question but also directly in the header.

Do not leave fields in the header blank. If the data is unknown, does not exist, or is not applicable, then write "NONE", "UNKNOWN", "DOES NOT EXIST" or "NOT APPLICABLE".

Likewise, do not skip the "Back-page Worksheet". Take the time to read the definition of *household* to the respondent and to fill out the roster member-by-member. If you cut corners by asking only, "How many members does the household have?", many respondents will miscount or apply the wrong definition of *household*. Completing the "Back-page Worksheet" improves data quality because it mimics the practice of Tanzania's NBS in the 2018 HBS. The accuracy of the scorecard's estimates depends on the quality of recorded responses and especially strongly on the count of household members. Working through the "Back-page Worksheet" gives the best count.

Throughout the interview, apply the instructions in the "Interview Guide". Enumerators must be thoroughly trained on the "Guide" before they do any interviews, and they should carry a copy of the "Guide" with them to each interview.[10] Even though the scorecard is less difficult than other poverty-assessment tools, training and explicit definitions of the scorecard's terms and concepts are still essential.[11] Enumerators must scrupulously study and follow the "Guide".

Finally, on-going quality-control audits are wise if a program or its field agents collect their own data and if they believe there is an incentive to exaggerate poverty estimates (for example, if they expect to be rewarded for higher poverty rates).[12]

---

[10] The "Interview Guide" is the only source of guidance for enumerators. All other issues of interpretation should be left to the judgment of enumerators and respondents, as this seems to be what Tanzania's NBS did in the 2018 HBS.

[11] Merely reading through the scorecard with enumerators is not adequate training.

[12] Matul and Kline, 2003. If a program does not want enumerators or respondents to know the scorecard's points, then it can use a mobile data-collection app or a paper version of the scorecard that omits the points, with scores computed later at an office. Even if points are hidden, however, enumerators and respondents can use common sense to guess how responses are linked with poverty.

Figure 3: First example household, filled-in scorecard

			Name	Identifier
Interview ID:	A123			
Interview date:	13JUN2020	Direct participant:	ANNA JACKSON	1V0276FZ7
Country:	TZA	Field agent:	UNKNOWN	UNKNOWN
Scorecard:	003	Service point:	NORTHWEST CLINIC	NWC
Sampling weight:	UNKNOWN	Number of household members:	NINE	

Question	Response	Points	
1. In which region does the household live?	A. Ruvuma, or Lindi	0	
	B. Dodoma, Mwanza, Rukwa, or Shinyanga	4	4
	C. Tanga, Morogoro, Arusha, or Iringa	6	
	D. Pwani, Geita, Simiyu, Manyara, Njombe, or Katavi	7	
	E. Kigoma, Mtwara, Singida, Tabora, or Songwe	8	
	F. Dar es Salaam, Kagera, Mbeya, Kilimanjaro, or Mara	12	
2. How many members does the household have? (*Based on the "Back-page Worksheet"*)	A. Eight or more	0	0
	B. Seven	5	
	C. Six	6	
	D. Five	9	
	E. Four	13	
	F. Three	19	
	G. Two	28	
	H. One	40	
3. What is the main building material of the floor?	A. Earth/sand, dung, vinyl or asphalt strips, wood planks, palm/bamboo, or other	0	0
	B. Cement, ceramic tiles, or parquet or polished wood	6	
4. Does the household own any charcoal stoves?	A. No	0	
	B. Yes	3	3
5. Does the household own any tables?	A. No	0	0
	B. Yes	3	
6. Does any member of the household own a television?	A. No	0	0
	B. Yes	12	
7. Does any member of the household own a mobile phone?	A. No	0	0
	B. Yes	6	
8. Does any member of the household own any cattle, goats, sheep, or hogs?	A. No	0	
	B. Yes	3	3
9. In the past week, did the household consume any chicken, goat, fish, beef, or other meat?	A. No	0	
	B. Yes	7	7
10. How many meals does the household usually have per day?	A. Two or less	0	0
	B. Three or more	8	

scorocs.com

Score: 4 + 0 + 0 + 3 + 0 + 0 + 0 + 3 + 7 + 0 = 17

Figure 4: First example household, filled-in "Back-page Worksheet"

First name or nickname?
1. ANNA
2. BILLY
3. CHARLES
4. DARLA
5. EUGENE
6. FREDA
7. GRETA
8. HANK
9. IRIS
10. —
11. —
12. —
13. —
14. —
15. —
16. —
17. —
Number of household members: NINE

2.3 First example household

The points for the first example household's responses add up to a score of 17 (Figure 3 and Figure 4).

For a given poverty line, Figure 1 lists poverty likelihoods by score range. A score of 17 falls in the first range of 0–17. For 150% of the national poverty line, the poverty likelihood for scores of 0–17 is 94.8 percent. That is, the scorecard estimates that 94.8 percent of households in Tanzania with a score of 0–17 have consumption expenditure below 150% of the national line.

Figure 5: The first example household's score of 17 implies a poverty likelihood of 94.8 percent for 150% of the national line (excerpted from Figure 1)

	Poverty likelihood (%)			
		National		
Score	Food	100%	150%	200%
0–17	40.0	71.4	94.8	98.4
18–21	19.8	60.1	89.8	96.3
22–24	19.4	58.2	86.5	95.9
25–26	17.5	53.8	81.9	92.5
27–28	12.2	46.1	81.1	92.1
29–30	12.2	36.9	78.1	91.6
31–32	9.2	32.3	73.3	91.6
33–34	7.0	31.8	72.8	89.6
35–36	6.0	28.8	68.0	87.1
. . .	. . .	. . .	. . .	. . .

Figure 6: Second example household, filled-in scorecard

			Name	Identifier
Interview ID:	B456			
Interview date:	30JUN2020	Direct participant:	JOHN BROWN	2W3120ZG8
Country:	TZA	Field agent:	UNKNOWN	UNKNOWN
Scorecard:	003	Service point:	NORTHWEST CLINIC	NWC
Sampling weight:	UNKNOWN	Number of household members:	FIVE	

Question	Response	Points	
1. In which region does the household live?	A. Ruvuma, or Lindi	0	
	B. Dodoma, Mwanza, Rukwa, or Shinyanga	**4**	**4**
	C. Tanga, Morogoro, Arusha, or Iringa	6	
	D. Pwani, Geita, Simiyu, Manyara, Njombe, or Katavi	7	
	E. Kigoma, Mtwara, Singida, Tabora, or Songwe	8	
	F. Dar es Salaam, Kagera, Mbeya, Kilimanjaro, or Mara	12	
2. How many members does the household have? (*Based on the "Back-page Worksheet"*)	**A. Eight or more**	**0**	
	B. Seven	5	
	C. Six	6	
	D. Five	9	9
	E. Four	13	
	F. Three	19	
	G. Two	28	
	H. One	40	
3. What is the main building material of the floor?	**A. Earth/sand, dung, vinyl or asphalt strips, wood planks, palm/bamboo, or other**	**0**	**0**
	B. Cement, ceramic tiles, or parquet or polished wood	6	
4. Does the household own any charcoal stoves?	A. No	0	
	B. Yes	**3**	**3**
5. Does the household own any tables?	**A. No**	**0**	**0**
	B. Yes	3	3
6. Does any member of the household own a television?	**A. No**	**0**	**0**
	B. Yes	12	
7. Does any member of the household own a mobile phone?	**A. No**	**0**	
	B. Yes	6	0
8. Does any member of the household own any cattle, goats, sheep, or hogs?	A. No	0	0
	B. Yes	**3**	
9. In the past week, did the household consume any chicken, goat, fish, beef, or other meat?	A. No	0	
	B. Yes	**7**	**7**
10. How many meals does the household usually have per day?	**A. Two or less**	**0**	**0**
	B. Three or more	8	

scorocs.com

Score: 4 + 9 + 0 + 3 + 3 + 0 + 0 + 0 + 0 + 7 = 26

Figure 7: Second example household, filled-in "Back-page Worksheet"

First name or nickname?
1. JOHN
2. MARY
3. SUE
4. KIM
5. MONICA
6. —
7. —
8. —
9. —
10. —
11. —
12. —
13. —
14. —
15. —
16. —
17. —
Number of household members: FIVE

2.4 Second example household

The points for the second example household's responses add up to a score of 26 (Figure 6 and Figure 7).

In Figure 1, a score of 26 falls in the range of 25–26. For 150% of the national poverty line, the poverty likelihood for scores of 25–26 is 81.9 percent. The scorecard estimates that 81.9 percent of households in Tanzania with a score of 25–26 have consumption expenditure below 150% of the national line.

Figure 8: The second example household's score of 26 implies a poverty likelihood of 81.9 percent for 150% of the national line (excerpt from Figure 1)

	Poverty likelihood (%)			
		National		
Score	Food	100%	150%	200%
0–17	40.0	71.4	94.8	98.4
18–21	19.8	60.1	89.8	96.3
22–24	19.4	58.2	86.5	95.9
25–26	17.5	53.8	81.9	92.5
27–28	12.2	46.1	81.1	92.1
29–30	12.2	36.9	78.1	91.6
31–32	9.2	32.3	73.3	91.6
33–34	7.0	31.8	72.8	89.6
35–36	6.0	28.8	68.0	87.1
. . .	. . .	. . .	. . .	. . .

3. How to calculate scorecard estimates

This section tells how to estimate:

- Head-count poverty rates for a single time period for in-coming participants
- Net changes in poverty rates across two time periods for on-going participants

It also tells how to use these estimated poverty rates to estimate:

- Number of poor people in the households of in-coming participants
- Net number of poor people in the households of on-going participants who rose above a poverty line

3.1 Head-count poverty rates in a single time period

The *head-count poverty rate* is the share of people in participating households in which total household consumption expenditure (divided by the number of members in the household or by the number of adult equivalents in the household) is below a given poverty line.

An estimate of the head-count poverty rate is the household-size-weighted average of poverty likelihoods from a scored sample, adjusted for the scorecard's known estimation error.

To illustrate the calculation, suppose that a pro-poor program opens a new service point in rural Dodoma in 2020. In that calendar year, it enrolls 1,000 in-coming households, from which it scores a simple random sample[13] of two households.[14]

The program judges that 150% of the national poverty line is the most-relevant line for its purposes. For that line and for snapshot estimates of poverty rates in one period, the scorecard's known estimation error is –1.8 percentage points (Figure 2).

The first example household has nine members and is interviewed on June 13, 2020 (Figure 3 and Figure 4). With a score of 17, it has a poverty likelihood for 150% of the national line of 94.8 percent (Figure 1).

The second example household has five members and is interviewed on June 30, 2020 (Figure 6 and Figure 7). Its score of 26 corresponds with a poverty likelihood of 81.9 percent.

---

[13] In a *simple random sample*, all households in the population have the same selection probability. This paper does not discuss samples in which different households have different selection probabilities.

[14] Of course, estimates based on such an unrealistically small sample have wide margins of error, but a small sample facilitates the arithmetic in the examples.

The estimated head-count poverty rate for the population of in-coming households in the 2020 calendar-year cohort in rural Dodoma is the household-size-weighted average of the estimated poverty likelihoods of the sampled households, less the known estimation error. Expressing poverty likelihoods and the estimation error as proportions between 0 and 1 rather than percentages between 0 and 100, this is:

$$\frac{9 \cdot 0.948 + 5 \cdot 0.819}{9 + 5} - (-0.018) \approx \frac{12.63}{14} + 0.018 \approx 0.920 = 92.0 \text{ percent.}$$

In the nine in the "9 · 0.948" term in the numerator is the number of members (household size) in the first household, and 0.948 is the first household's estimated poverty likelihood as proportion.

In the same way, the five in the numerator's "5 · 0.819" is the number of members in the second household, and 0.819 is the second household's estimated poverty likelihood.

The "9 + 5" in the denominator is the sum of the weights—that is, the number of household members—across the two sampled households.

The "–0.018" is the scorecard's estimation error for this poverty line (Figure 2). Because unadjusted estimates tend to be too low by 1.8 percentage points, they are adjusted upwards by subtracting –1.8 (which is equivalent to adding 1.8). This is akin to how an archer whose arrows tend to miss a little to the left of the bulls-eye will adjust his/her aim to be a little to the right of the bulls-eye.

The estimated head-count poverty rate for the population is 92.0 percent. Again, this is the household-size-weighted average of the two sampled households' poverty likelihoods, adjusted for the known estimation error.[15]

With hundreds or thousands of interviewed households, the calculations are by an app or in a spreadsheet (Figure 9 below).

---

[15] Be careful; the estimated poverty rate is *not* the single poverty likelihood associated with the household-size-weighted average score, which here is (9·17 + 5·26) ÷ (9 + 5) ≈ 20. This average score of 20 corresponds to a poverty likelihood of 89.8 percent (Figure 1), giving an error-adjusted poverty rate of 89.8 – (–1.8) = 91.6 percent. This differs from the 92.0 percent found as the household-size-weighted average of the two individual likelihoods associated with each of the two scores. Unlike likelihoods, scores are ordinal symbols, like colors in the spectrum or syllables in a solfège scale. Because scores are ordinal, they cannot be added up nor averaged. Only three operations are valid for scores: conversion to likelihoods, analysis of distributions, or comparison with a cut-off for segmentation (Schreiner, 2012). In general, programs should analyze likelihoods, not scores.

Figure 9: Spreadsheet calculation to estimate the head-count poverty rate and number of poor people in a population of in-coming participants in a period

	A	B	C	D	E	F	G
1	Survey	Interview date	ID of direct participant	Number of household members	Score	Poverty likelihood (%)	Estimated number of poor household members
2	Baseline	13-Jun-20	1V0276FZ7	9	17	94.8	8.53 = (D2*F2)/100
3	Baseline	30-Jun-20	2W3120ZG8	5	26	81.9	4.10 = (D3*F3)/100
4			Sum:	14 = SUM(D2:D3)			12.63 = SUM(G2:G3)
5			Average:	7.0 = AVERAGE(D2:D3)			
6							
7	Estimated scorecard error for this poverty line (percentage points):						–1.8
8							
9				Estimated head-count poverty rate (%):			92.0 = (G4/D4)*100-G7
10							
11				Households in the population:			1,000
12							
13				People in households in the population:			7,000 = G11*D5
14							
15				Number of poor people in population:			6,440 = (G9/100)*G13
16	Rows of data are sorted by Round, then by Interview date, then by Direct participant ID.						

This snapshot estimate in a single time period tends to be more relevant for in-coming participants who joined in the current period than for on-going participants who joined in past periods. This is because fulfilling a pro-poor mission implies that some share of new participants be poor by some definition of *poverty*.[16] To be pro-poor, a bare-minimum standard is that the poverty rate of in-coming participants exceed that of the country as a whole or that of the area where the program works.

To help with benchmarking poverty-rate estimates, Figure 10 reports head-count poverty rates from the 2018 HBS for all 15 poverty lines by urban/rural/all for Tanzania overall and for each of the 26 regions defined by NBS for the HBS. In the example of rural Dodoma, the head-count poverty rate for 150% of the national line is 71.4 percent. Thus, the example program is pro-poor in the sense that its in-coming participants have an above-average estimated poverty rate (92.0 percent).

The text that illustrates the calculation of the scorecard estimate of the number of poor people in a single time period follows after Figure 10, which stretches across the next eight pages. The regions in Figure 10 begin with all-Tanzania overall, followed in alphabetical order by the 26 regions defined by the NBS for the 2018 HBS. The HBS does not cover Zanzibar.

[16] The scorecard uses a consumption-based definition of *poverty*. Common non-consumption definitions include: being rural, agricultural, landless, or unemployed; living in a given region; having a head who is illiterate, female, or an ethnic minority; or having a member who is pregnant, handicapped, elderly, or young.

Figure 10: (Tanzania overall, Arusha, and Dar es Salaam): Poverty lines and head-count poverty rates by urban/rural/all in 2018

	Region/ Area	Line or Rate	*n*	Poverty lines and poverty rates														
				National				Intl. 2011 PPP				Percentile-based lines						
				Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
Tanzania	Urban	Line	2,788	1,134	1,657	2,486	3,315	1,837	3,093	5,317	20,977	953	1,183	1,605	1,870	2,193	3,306	4,669
		Rate		4.4	15.8	34.4	52.7	29.3	59.1	83.4	99.2	5.2	10.8	23.2	30.2	39.6	64.0	79.7
	Rural	Line	6,674	1,095	1,600	2,401	3,201	1,774	2,987	5,134	20,256	920	1,142	1,550	1,806	2,118	3,192	4,509
		Rate		9.7	31.3	63.2	80.2	58.1	85.4	96.3	99.9	12.3	24.2	47.8	59.2	69.5	87.5	94.8
	All	Line	9,462	1,108	1,619	2,428	3,237	1,794	3,021	5,192	20,486	931	1,155	1,568	1,827	2,142	3,228	4,560
		Rate		8.0	26.4	54.0	71.5	48.9	77.0	92.2	99.7	10.0	19.9	40.0	50.0	60.0	80.0	90.0
Arusha	Urban	Line	132	1,209	1,767	2,650	3,533	1,958	3,297	5,667	22,360	1,016	1,261	1,711	1,994	2,338	3,524	4,977
		Rate		1.7	11.6	37.0	51.5	30.8	57.3	88.4	98.1	1.7	4.0	20.4	31.0	38.0	66.2	85.9
	Rural	Line	271	1,209	1,767	2,650	3,534	1,958	3,298	5,668	22,364	1,016	1,261	1,712	1,994	2,338	3,524	4,978
		Rate		10.1	30.2	64.8	81.3	59.7	86.0	96.9	99.8	14.2	25.2	49.9	61.0	73.1	88.2	96.3
	All	Line	403	1,209	1,767	2,650	3,534	1,958	3,298	5,668	22,363	1,016	1,261	1,712	1,994	2,338	3,524	4,978
		Rate		7.6	24.7	56.5	72.4	51.1	77.4	94.4	99.3	10.5	18.9	41.1	52.1	62.6	81.6	93.2
Dar es Salaam	Urban	Line	797	1,212	1,771	2,657	3,543	1,963	3,306	5,682	22,420	1,019	1,264	1,716	1,999	2,344	3,533	4,990
		Rate		2.3	8.0	20.0	39.5	16.0	45.8	72.8	98.8	2.7	5.6	12.7	16.4	23.2	49.9	67.4
	Rural	Line	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
		Rate		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	All	Line	797	1,212	1,771	2,657	3,543	1,963	3,306	5,682	22,420	1,019	1,264	1,716	1,999	2,344	3,533	4,990
		Rate		2.3	8.0	20.0	39.5	16.0	45.8	72.8	98.8	2.7	5.6	12.7	16.4	23.2	49.9	67.4

Source: 2018 HBS.

Poverty rates are percentages.

National poverty lines are TZA per adult equivalent, per day.

Intl. 2011 PPP poverty lines and percentile-based lines are TZA per-person, per-day.

All poverty lines are TZA in prices in Tanzania as a whole on average during the 2018 HBS fieldwork.

Figure 10: (Dodoma, Geita, and Iringa): Poverty lines and head-count poverty rates by urban/rural/all in 2018

Region/ Area	Line or Rate	*n*	Poverty lines and poverty rates														
			National				Intl. 2011 PPP				Percentile-based lines						
			Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
Dodoma	Urban	Line	1,137	1,662	2,492	3,323	1,841	3,101	5,330	21,029	955	1,186	1,610	1,875	2,199	3,314	4,681
		Rate	0.0	6.7	19.1	39.4	10.8	39.4	72.0	100.0	0.0	0.0	8.3	10.8	22.4	53.2	69.5
		n	46														
	Rural	Line	1,137	1,661	2,491	3,322	1,841	3,100	5,328	21,022	955	1,185	1,609	1,874	2,198	3,313	4,679
		Rate	4.8	28.0	71.4	86.2	63.7	87.9	97.4	100.0	8.6	20.4	49.5	64.7	76.3	88.6	95.8
		n	287														
	All	Line	1,137	1,661	2,492	3,322	1,841	3,100	5,329	21,023	955	1,185	1,609	1,875	2,198	3,313	4,680
		Rate	3.7	23.2	59.6	75.7	51.8	77.0	91.7	100.0	6.6	15.8	40.2	52.5	64.2	80.6	89.8
		n	333														
Geita	Urban	Line	1,091	1,594	2,391	3,189	1,767	2,976	5,114	20,178	917	1,138	1,544	1,799	2,110	3,180	4,491
		Rate	10.8	32.4	55.8	67.3	50.2	76.4	95.4	100.0	15.3	22.5	43.1	52.3	61.5	82.3	92.6
		n	70														
	Rural	Line	1,090	1,593	2,390	3,186	1,766	2,973	5,111	20,164	916	1,137	1,543	1,798	2,108	3,178	4,488
		Rate	15.2	38.4	73.3	85.9	69.8	91.6	99.3	100.0	19.0	32.2	58.8	70.1	76.5	93.0	98.1
		n	331														
	All	Line	1,090	1,593	2,390	3,187	1,766	2,974	5,111	20,166	916	1,137	1,543	1,798	2,109	3,178	4,489
		Rate	14.5	37.5	70.6	83.1	66.8	89.2	98.7	100.0	18.4	30.6	56.3	67.3	74.2	91.3	97.2
		n	401														
Iringa	Urban	Line	1,141	1,667	2,501	3,334	1,847	3,111	5,348	21,099	959	1,190	1,615	1,881	2,206	3,325	4,696
		Rate	0.7	7.6	25.9	66.5	27.1	72.0	95.1	100.0	0.7	5.6	18.7	30.6	45.5	73.9	88.7
		n	84														
	Rural	Line	1,142	1,670	2,504	3,339	1,850	3,116	5,356	21,131	960	1,191	1,617	1,884	2,210	3,330	4,704
		Rate	11.4	31.5	64.4	82.3	60.8	88.4	97.2	100.0	11.9	26.2	49.8	62.3	72.9	89.5	96.1
		n	215														
	All	Line	1,142	1,669	2,503	3,338	1,849	3,115	5,353	21,121	960	1,191	1,617	1,883	2,208	3,329	4,701
		Rate	8.0	24.0	52.2	77.3	50.1	83.2	96.6	100.0	8.4	19.6	39.9	52.2	64.2	84.6	93.7
		n	299														

Source: 2018 HBS.

Poverty rates are percentages.

National poverty lines are TZA per adult equivalent, per day.

Intl. 2011 PPP poverty lines and percentile-based lines are TZA per-person, per-day.

All poverty lines are TZA in prices in Tanzania as a whole on average during the 2018 HBS fieldwork.

Figure 10: (Kagera, Katavi, and Kigoma): Poverty lines and head-count poverty rates by urban/rural/all in 2018

	Region/ Area	Line or Rate	*n*	Poverty lines and poverty rates														
				National				Intl. 2011 PPP				Percentile-based lines						
				Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
Kagera	Urban	Line	48	968	1,415	2,123	2,831	1,568	2,642	4,540	17,913	814	1,010	1,371	1,597	1,873	2,823	3,987
		Rate		0.0	2.9	11.6	38.8	11.6	45.8	84.1	100.0	0.0	2.9	5.9	11.6	16.7	49.2	78.5
	Rural	Line	360	969	1,417	2,125	2,833	1,570	2,644	4,544	17,930	815	1,011	1,372	1,599	1,875	2,826	3,991
		Rate		12.8	33.8	63.9	81.9	57.3	85.8	96.9	100.0	16.7	28.3	49.1	58.2	72.7	87.0	94.8
	All	Line	408	969	1,417	2,125	2,833	1,570	2,644	4,544	17,929	815	1,011	1,372	1,599	1,875	2,825	3,991
		Rate		12.0	31.9	60.6	79.2	54.4	83.3	96.1	100.0	15.6	26.7	46.4	55.2	69.2	84.7	93.8
Katavi	Urban	Line	84	1,059	1,547	2,321	3,095	1,715	2,888	4,963	19,583	890	1,104	1,499	1,746	2,048	3,086	4,359
		Rate		2.0	28.2	61.7	72.0	49.0	85.9	94.6	100.0	7.0	16.0	42.1	50.6	64.7	86.7	92.6
	Rural	Line	215	1,059	1,548	2,321	3,095	1,715	2,889	4,965	19,588	890	1,104	1,499	1,747	2,048	3,087	4,360
		Rate		10.7	29.1	60.6	82.3	56.1	89.4	98.5	100.0	12.7	24.8	45.3	58.2	71.2	90.6	97.6
	All	Line	299	1,059	1,548	2,321	3,095	1,715	2,888	4,965	19,587	890	1,104	1,499	1,746	2,048	3,087	4,360
		Rate		9.2	29.0	60.8	80.5	54.9	88.7	97.8	100.0	11.7	23.3	44.8	56.8	70.0	89.9	96.7
Kigoma	Urban	Line	60	950	1,388	2,082	2,776	1,538	2,591	4,453	17,568	798	991	1,345	1,566	1,837	2,769	3,910
		Rate		6.0	18.6	65.3	81.7	57.9	79.1	96.6	97.6	6.7	23.5	40.8	57.9	70.7	88.5	94.5
	Rural	Line	226	955	1,396	2,094	2,792	1,547	2,606	4,479	17,671	803	996	1,353	1,576	1,848	2,785	3,933
		Rate		18.1	42.5	67.3	84.8	61.7	89.4	97.2	100.0	20.2	33.1	56.6	64.3	72.8	90.9	97.2
	All	Line	286	953	1,393	2,090	2,787	1,544	2,601	4,470	17,636	801	994	1,350	1,572	1,844	2,779	3,926
		Rate		13.9	34.3	66.6	83.7	60.4	85.9	97.0	99.2	15.6	29.8	51.2	62.1	72.1	90.1	96.2

Source: 2018 HBS.

Poverty rates are percentages.

National poverty lines are TZA per adult equivalent, per day.

Intl. 2011 PPP poverty lines and percentile-based lines are TZA per-person, per-day.

All poverty lines are TZA in prices in Tanzania as a whole on average during the 2018 HBS fieldwork.

Figure 10: (Kilimanjaro, Lindi, and Manyara): Poverty lines and head-count poverty rates by urban/rural/all in 2018

Region/ Area	Line or Rate	*n*	Poverty lines and poverty rates														
			National				Intl. 2011 PPP				Percentile-based lines						
			Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
Kilimanjaro	Urban	Line	1,096	1,602	2,402	3,203	1,775	2,989	5,138	20,271	921	1,143	1,552	1,807	2,120	3,195	4,512
		Rate	5.6	11.5	35.9	51.6	29.8	61.9	86.5	97.4	4.3	10.0	16.2	29.8	38.3	64.3	79.5
		n	96														
	Rural	Line	1,096	1,602	2,403	3,204	1,775	2,990	5,139	20,276	921	1,143	1,552	1,808	2,120	3,195	4,513
		Rate	1.2	10.3	32.2	56.2	25.6	63.6	83.1	99.7	1.8	3.6	18.8	26.4	35.3	65.2	80.0
		n	312														
	All	Line	1,096	1,602	2,403	3,204	1,775	2,990	5,139	20,275	921	1,143	1,552	1,808	2,120	3,195	4,513
		Rate	2.1	10.5	33.0	55.3	26.5	63.3	83.7	99.2	2.3	4.9	18.3	27.1	35.9	65.0	79.9
		n	408														
Lindi	Urban	Line	1,174	1,716	2,574	3,431	1,901	3,202	5,504	21,715	987	1,224	1,662	1,936	2,271	3,422	4,833
		Rate	5.1	27.2	58.6	72.6	55.9	74.1	92.9	100.0	2.4	18.3	44.9	59.4	65.2	76.8	83.5
		n	59														
	Rural	Line	1,174	1,716	2,573	3,431	1,901	3,202	5,504	21,714	986	1,224	1,662	1,936	2,271	3,422	4,833
		Rate	18.0	40.9	73.9	85.5	68.3	89.7	97.2	100.0	20.4	30.9	58.5	69.2	75.6	89.8	96.7
		n	239														
	All	Line	1,174	1,716	2,573	3,431	1,901	3,202	5,504	21,714	986	1,224	1,662	1,936	2,271	3,422	4,833
		Rate	15.3	38.0	70.7	82.8	65.7	86.4	96.3	100.0	16.6	28.2	55.7	67.2	73.4	87.1	93.9
		n	298														
Manyara	Urban	Line	1,115	1,629	2,444	3,258	1,805	3,041	5,226	20,619	937	1,163	1,578	1,838	2,156	3,249	4,589
		Rate	0.0	3.1	3.1	21.8	3.1	46.6	85.4	97.2	0.0	3.1	3.1	3.1	7.3	56.5	80.2
		n	47														
	Rural	Line	1,115	1,629	2,444	3,258	1,805	3,041	5,226	20,620	937	1,163	1,578	1,839	2,156	3,250	4,590
		Rate	9.0	33.8	66.9	81.1	61.3	87.4	97.8	100.0	10.1	26.0	48.8	62.2	70.8	87.6	97.3
		n	249														
	All	Line	1,115	1,629	2,444	3,258	1,805	3,041	5,226	20,620	937	1,163	1,578	1,839	2,156	3,250	4,590
		Rate	8.0	30.5	60.1	74.8	55.1	83.0	96.4	99.7	9.1	23.6	43.9	55.9	64.0	84.3	95.5
		n	296														

Source: 2018 HBS.

Poverty rates are percentages.

National poverty lines are TZA per adult equivalent, per day.

Intl. 2011 PPP poverty lines and percentile-based lines are TZA per-person, per-day.

All poverty lines are TZA in prices in Tanzania as a whole on average during the 2018 HBS fieldwork.

Figure 10: (Mara, Mbeya, and Morogoro): Poverty lines and head-count poverty rates by urban/rural/all in 2018

Region/ Area	Line or Rate	*n*	Poverty lines and poverty rates														
			National				Intl. 2011 PPP				Percentile-based lines						
			Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
Mara	Urban	Line	1,096	1,602	2,403	3,204	1,775	2,990	5,140	20,278	921	1,143	1,552	1,808	2,120	3,196	4,514
		Rate	4.6	18.5	31.9	61.0	30.8	65.1	88.5	100.0	9.0	16.6	24.5	31.2	48.8	73.9	84.1
		n	72														
	Rural	Line	1,096	1,602	2,403	3,204	1,776	2,990	5,140	20,278	921	1,143	1,552	1,808	2,120	3,196	4,514
		Rate	1.6	24.4	61.4	83.1	57.5	86.2	97.0	100.0	6.2	17.1	44.8	59.5	70.0	88.2	94.6
		n	336														
	All	Line	1,096	1,602	2,403	3,204	1,775	2,990	5,140	20,278	921	1,143	1,552	1,808	2,120	3,196	4,514
		Rate	2.2	23.2	55.5	78.6	52.2	82.0	95.3	100.0	6.7	17.0	40.7	53.8	65.8	85.4	92.5
		n	408														
Mbeya	Urban	Line	1,095	1,600	2,401	3,201	1,774	2,987	5,134	20,257	920	1,142	1,550	1,806	2,118	3,192	4,509
		Rate	11.3	26.6	48.7	64.8	40.1	72.9	92.6	99.5	11.3	18.9	34.2	40.3	54.6	76.3	91.3
		n	156														
	Rural	Line	1,095	1,600	2,399	3,199	1,773	2,986	5,131	20,246	920	1,142	1,550	1,805	2,117	3,191	4,506
		Rate	4.8	17.2	35.0	57.8	31.9	70.0	91.9	99.9	5.1	9.9	25.5	34.1	43.9	75.5	89.5
		n	251														
	All	Line	1,095	1,600	2,400	3,200	1,773	2,986	5,133	20,251	920	1,142	1,550	1,806	2,118	3,191	4,508
		Rate	7.7	21.4	41.1	60.9	35.6	71.3	92.2	99.7	7.9	13.9	29.4	36.9	48.8	75.8	90.3
		n	407														
Morogoro	Urban	Line	1,115	1,629	2,444	3,258	1,805	3,041	5,226	20,620	937	1,163	1,578	1,839	2,156	3,250	4,590
		Rate	0.9	15.9	30.4	46.8	24.1	56.6	84.4	99.7	1.4	5.3	21.2	24.1	36.7	61.7	84.2
		n	114														
	Rural	Line	1,116	1,631	2,447	3,262	1,808	3,044	5,232	20,644	938	1,164	1,580	1,841	2,159	3,253	4,595
		Rate	7.0	22.0	58.6	78.5	49.0	84.1	95.4	99.7	5.2	15.2	37.9	50.8	62.3	85.4	92.4
		n	284														
	All	Line	1,116	1,630	2,445	3,260	1,807	3,043	5,230	20,633	937	1,163	1,579	1,840	2,157	3,252	4,593
		Rate	4.3	19.3	46.1	64.5	38.0	72.0	90.5	99.7	3.5	10.8	30.5	39.0	51.0	75.0	88.8
		n	398														

Source: 2018 HBS.

Poverty rates are percentages.

National poverty lines are TZA per adult equivalent, per day.

Intl. 2011 PPP poverty lines and percentile-based lines are TZA per-person, per-day.

All poverty lines are TZA in prices in Tanzania as a whole on average during the 2018 HBS fieldwork.

Figure 10: (Mtwara, Mwanza, and Njombe): Poverty lines and head-count poverty rates by urban/rural/all in 2018

	Region/ Area	Line or Rate	*n*	Poverty lines and poverty rates														
				National				Intl. 2011 PPP				Percentile-based lines						
				Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
Mtwara	Urban	Line	60	1,157	1,691	2,536	3,381	1,874	3,156	5,424	21,398	972	1,207	1,638	1,908	2,238	3,372	4,763
		Rate		1.8	23.6	29.4	43.2	28.2	59.0	87.6	100.0	2.4	3.2	24.3	28.2	33.0	68.3	85.7
	Rural	Line	240	1,160	1,695	2,542	3,390	1,878	3,163	5,437	21,450	974	1,209	1,642	1,913	2,243	3,380	4,774
		Rate		13.9	33.0	63.7	78.9	58.0	82.9	95.7	99.8	13.2	22.8	46.7	58.3	67.6	85.7	94.4
	All	Line	300	1,159	1,693	2,540	3,386	1,876	3,160	5,431	21,429	974	1,208	1,640	1,911	2,241	3,377	4,770
		Rate		9.0	29.1	49.7	64.3	45.8	73.1	92.4	99.9	8.8	14.8	37.6	46.0	53.5	78.6	90.9
Mwanza	Urban	Line	156	1,084	1,585	2,377	3,170	1,756	2,958	5,084	20,058	911	1,131	1,535	1,788	2,097	3,161	4,465
		Rate		8.2	29.8	54.2	74.1	51.0	78.0	93.2	99.9	10.9	22.6	42.2	51.8	62.6	79.7	92.4
	Rural	Line	252	1,084	1,584	2,376	3,168	1,755	2,956	5,081	20,047	911	1,130	1,534	1,787	2,096	3,159	4,462
		Rate		10.1	37.3	71.1	86.5	68.8	89.6	97.0	100.0	15.2	29.1	56.1	68.8	79.8	91.3	96.1
	All	Line	408	1,084	1,584	2,376	3,168	1,756	2,957	5,082	20,051	911	1,131	1,535	1,788	2,097	3,160	4,463
		Rate		9.4	34.6	65.0	82.1	62.4	85.4	95.6	100.0	13.7	26.8	51.1	62.7	73.6	87.1	94.8
Njombe	Urban	Line	72	1,102	1,611	2,416	3,221	1,785	3,006	5,167	20,385	926	1,149	1,560	1,818	2,132	3,213	4,537
		Rate		4.7	8.0	26.1	48.4	20.5	58.0	85.0	100.0	4.7	4.7	14.2	21.8	27.5	58.9	78.8
	Rural	Line	228	1,103	1,612	2,418	3,225	1,787	3,009	5,172	20,406	927	1,151	1,562	1,820	2,134	3,216	4,542
		Rate		2.5	16.3	56.1	76.2	44.6	81.3	95.1	99.8	2.5	12.5	33.0	51.2	65.7	83.7	93.3
	All	Line	300	1,103	1,612	2,417	3,223	1,786	3,008	5,170	20,398	927	1,150	1,561	1,819	2,133	3,215	4,540
		Rate		3.3	13.2	44.8	65.7	35.5	72.4	91.3	99.9	3.3	9.5	25.9	40.1	51.3	74.3	87.8

Source: 2018 HBS.

Poverty rates are percentages.

National poverty lines are TZA per adult equivalent, per day.

Intl. 2011 PPP poverty lines and percentile-based lines are TZA per-person, per-day.

All poverty lines are TZA in prices in Tanzania as a whole on average during the 2018 HBS fieldwork.

Figure 10: (Shinyanga, Simiyu, and Singida): Poverty lines and head-count poverty rates by urban/rural/all in 2018

Region/ Area	Line or Rate	*n*	Poverty lines and poverty rates														
			National				Intl. 2011 PPP				Percentile-based lines						
			Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
Shinyanga	Urban	Line	1,144	1,672	2,508	3,344	1,853	3,120	5,363	21,160	961	1,193	1,620	1,887	2,213	3,335	4,710
		Rate	6.7	18.2	52.6	68.1	41.7	77.6	91.8	99.1	3.0	11.9	27.6	43.6	59.0	78.9	88.1
		n	96														
	Rural	Line	1,142	1,669	2,504	3,338	1,850	3,115	5,354	21,126	960	1,191	1,617	1,884	2,209	3,329	4,702
		Rate	8.4	33.8	74.5	89.6	68.2	94.3	99.9	100.0	12.0	26.7	58.7	69.7	79.5	95.5	99.7
		n	311														
	All	Line	1,142	1,669	2,504	3,339	1,850	3,116	5,356	21,130	960	1,191	1,617	1,884	2,209	3,330	4,703
		Rate	8.2	31.9	71.9	87.0	65.0	92.3	98.9	99.9	10.9	24.9	54.9	66.5	77.0	93.5	98.3
		n	407														
Simiyu	Urban	Line	1,104	1,613	2,420	3,227	1,788	3,011	5,176	20,420	928	1,151	1,563	1,821	2,135	3,218	4,545
		Rate	3.4	7.0	35.6	47.0	29.6	66.1	92.6	97.2	4.5	9.5	21.7	30.9	42.6	66.7	85.9
		n	48														
	Rural	Line	1,103	1,612	2,418	3,224	1,787	3,009	5,172	20,404	927	1,150	1,562	1,819	2,134	3,216	4,542
		Rate	7.8	42.1	74.1	88.4	69.9	92.5	98.1	99.7	15.2	32.1	58.6	70.7	81.3	94.2	98.1
		n	356														
	All	Line	1,103	1,612	2,418	3,224	1,787	3,009	5,172	20,406	927	1,151	1,562	1,819	2,134	3,216	4,542
		Rate	7.5	39.2	71.0	85.0	66.6	90.3	97.6	99.5	14.4	30.3	55.6	67.4	78.1	91.9	97.1
		n	404														
Singida	Urban	Line	1,128	1,649	2,473	3,297	1,827	3,077	5,289	20,866	948	1,177	1,597	1,860	2,182	3,288	4,644
		Rate	0.0	20.1	44.1	59.9	38.2	69.5	94.0	100.0	5.5	8.5	29.9	40.1	47.4	70.7	92.8
		n	36														
	Rural	Line	1,126	1,646	2,468	3,291	1,824	3,071	5,279	20,827	946	1,174	1,594	1,857	2,178	3,282	4,636
		Rate	9.8	34.7	67.4	84.2	66.9	89.9	99.4	100.0	13.1	27.5	55.2	67.9	75.3	92.3	97.9
		n	263														
	All	Line	1,126	1,646	2,469	3,291	1,824	3,072	5,279	20,829	946	1,174	1,594	1,857	2,178	3,283	4,636
		Rate	9.3	34.0	66.3	83.0	65.5	88.9	99.1	100.0	12.7	26.6	54.0	66.5	74.0	91.3	97.6
		n	299														

Source: 2018 HBS.

Poverty rates are percentages.

National poverty lines are TZA per adult equivalent, per day.

Intl. 2011 PPP poverty lines and percentile-based lines are TZA per-person, per-day.

All poverty lines are TZA in prices in Tanzania as a whole on average during the 2018 HBS fieldwork.

Figure 10: (Songwe, Tabora, and Tanga): Poverty lines and head-count poverty rates by urban/rural/all in 2018

Region/ Area		Line or Rate	*n*	Poverty lines and poverty rates														
				National				Intl. 2011 PPP				Percentile-based lines						
				Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
Songwe	Urban	Line	72	1,070	1,564	2,346	3,127	1,733	2,919	5,016	19,791	899	1,116	1,515	1,765	2,069	3,119	4,405
		Rate		0.0	9.4	40.8	61.1	24.4	63.2	84.8	99.5	0.0	0.0	18.5	31.5	44.2	69.8	83.5
	Rural	Line	227	1,070	1,564	2,345	3,127	1,733	2,918	5,016	19,789	899	1,116	1,515	1,764	2,069	3,119	4,405
		Rate		7.5	24.4	57.4	75.3	50.8	83.6	95.8	100.0	9.3	17.7	37.8	51.3	65.2	86.1	94.6
	All	Line	299	1,070	1,564	2,345	3,127	1,733	2,918	5,016	19,790	899	1,116	1,515	1,764	2,069	3,119	4,405
		Rate		5.6	20.7	53.3	71.7	44.2	78.5	93.1	99.9	7.0	13.3	33.0	46.4	60.0	82.0	91.8
Tabora	Urban	Line	48	1,043	1,525	2,287	3,049	1,690	2,846	4,891	19,297	877	1,088	1,477	1,721	2,018	3,041	4,295
		Rate		14.2	27.7	40.6	57.2	39.2	64.0	85.0	100.0	18.6	30.0	32.9	39.2	48.1	69.5	76.7
	Rural	Line	252	1,042	1,523	2,285	3,047	1,688	2,843	4,887	19,280	876	1,087	1,476	1,719	2,016	3,038	4,291
		Rate		10.6	35.7	66.3	82.4	64.4	89.6	96.0	99.3	12.8	29.4	53.6	65.6	74.3	91.1	94.8
	All	Line	300	1,042	1,524	2,285	3,047	1,688	2,844	4,887	19,283	876	1,087	1,476	1,719	2,016	3,039	4,292
		Rate		11.1	34.5	62.5	78.7	60.7	85.8	94.3	99.4	13.7	29.5	50.5	61.7	70.4	87.9	92.1
Tanga	Urban	Line	84	1,107	1,619	2,428	3,237	1,794	3,021	5,192	20,485	931	1,155	1,568	1,826	2,142	3,228	4,560
		Rate		4.0	8.9	32.3	48.5	18.9	55.2	75.5	99.0	4.0	6.9	18.0	22.0	33.0	57.3	74.7
	Rural	Line	322	1,108	1,619	2,429	3,238	1,794	3,022	5,194	20,494	931	1,156	1,569	1,827	2,143	3,230	4,562
		Rate		8.9	23.6	53.0	71.7	47.0	77.1	95.3	99.9	9.4	18.5	38.0	47.6	58.8	83.3	92.8
	All	Line	406	1,108	1,619	2,429	3,238	1,794	3,022	5,194	20,492	931	1,155	1,568	1,827	2,143	3,229	4,561
		Rate		8.0	21.0	49.3	67.5	42.0	73.2	91.7	99.7	8.4	16.4	34.4	43.0	54.1	78.7	89.5

Source: 2018 HBS.

Poverty rates are percentages.

National poverty lines are TZA per adult equivalent, per day.

Intl. 2011 PPP poverty lines and percentile-based lines are TZA per-person, per-day.

All poverty lines are TZA in prices in Tanzania as a whole on average during the 2018 HBS fieldwork.

3.2 Number of poor people in a single time period

Fulfilling a pro-poor mission depends not only on the *poverty rate* of in-coming participants but also on the *number* of poor in-coming participants. After all, a smaller program whose few participants have a higher poverty rate may serve fewer poor people than a larger program whose many participants have a lower poverty rate.[17]

The first step in estimating the number of poor people in one period is to estimate the number of household members in the population of in-coming households. In our two-household example with simple random sampling, this is the equal-weighted average of the number of people in the sampled households:

$$\frac{9+5}{1+1} = \frac{14}{2} = 7.0 \text{ people.}$$

The second step is to estimate the total number of people in the population of in-coming households. The example program has 1,000 in-coming households in its first calendar-year, each with an estimated 7.0 members. The estimated number of people in the households of in-coming participants is then 1,000 · 7.0 = 7,000.

The third and final step is to multiply the estimated poverty rate (here, 92.0 percent, or 0.920) by the estimated number of people in in-coming households (here, 7,000). This gives 7,000 · 0.920 = 6,440 people (Figure 9).

All else constant, the *number* of in-coming participants who are poor is more important than the *share* of in-coming participants who are poor. Both estimates are useful,[18] but increasing the share who are poor is only a means to the end of increasing the number who are poor.

In turn, increasing the number of in-coming participants who are poor is only a means to the end of increasing the net number of on-going participants who rise above a poverty line.

[17] Navajas *et al*. (2000).

[18] Schreiner (2014) tells how to report and analyze scorecard estimates.

3.3 Net changes in poverty rates across two time periods for on-going participants

The estimated net change in a population's poverty rate is the difference between estimated poverty rates at follow-up versus baseline.

After baseline, two sampling approaches are possible for the follow-up round:

- *One sample scored twice*: Score the same sample that was scored at baseline
- *Two independent samples*: Score a new sample from the same population that was scored at baseline

Given the scorecard's assumptions, both approaches are unbiased, but scoring one sample twice has smaller margins of error than does scoring two independent samples.

3.3.1 Annual net change in poverty rates with one sample scored twice

When the follow-up sample is made up of the same households as the baseline sample,[19] then the estimated annual net change in the poverty rate of the population of on-going participants is the average-household-size-weighted average of the change in each scored household's poverty likelihood, divided by the household-size-weighted average of the years between each household's interviews.[20]

Continuing the earlier example, suppose that the first household at follow-up has eight members (rather than nine as at baseline) and is scored a second time on August 13, 2023, which is 1,156 days (about 3.17 years) after its first interview on June 13, 2020. Its score is now 22 (rather than 17), so its poverty likelihood for 150% of the national line has decreased from 94.8 to 86.5 percent (Figure 1).

Suppose also that the second household now has four members (rather than five as at baseline) and is scored a second time on May 15, 2023, which is 1,049 days (about 2.87 years) after its first interview on June 30, 2020. Its score is now 27 (rather than 26), so its poverty likelihood has decreased from 81.9 to 81.1percent.

[19] Or when the follow-up sample is a random sample of the baseline sample.
[20] Estimates of change do not directly adjust for the estimation error in snapshot estimates because—given the scorecard's assumptions—this error washes out when comparing follow-up with baseline. Error due to divergence from assumptions is unknown, and there is no direct way to adjust for it.

With poverty likelihoods expressed as proportions between 0 and 1, the average-household-size-weighted average of the change in each scored household's poverty likelihood is –5.7 percentage points:

$$\frac{\left(\frac{9+8}{2}\right)\cdot(0.865-0.948)+\left(\frac{5+4}{2}\right)\cdot(0.811-0.819)}{\left(\frac{9+8}{2}\right)+\left(\frac{5+4}{2}\right)}\approx\frac{-0.706-0.036}{13}\approx-0.057.$$

The head-count poverty rate decreased (improved) by 5.7 percentage points (not by 5.7 percent) between baseline and follow-up.

For clarity—and because the time between interviews varies across scored households—this estimate should be annualized by dividing it by the average-household-size-weighted average of years between the two interviews:

$$\frac{\left(\frac{9+8}{2}\right)\cdot 3.17+\left(\frac{5+4}{2}\right)\cdot 2.87}{\left(\frac{9+8}{2}\right)+\left(\frac{5+4}{2}\right)}\approx\frac{26.95+12.92}{13}\approx 3.07 \text{ years.}$$

The annual, non-compounded rate of net change is then the percentage-point change in the poverty rate, divided by the average years between interviews: –5.7 ÷ 3.07 ≈ –1.9 percentage points per year.[21] The negative change means that poverty decreased.[22]

In practice, the calculations are done with an app or spreadsheet (Figure 11 below).

[21] *Percentage points* are distinct from *percentages* (or *percents*). On the one hand, if the baseline poverty rate is 50.0 percent, and if there is a *10.0-percent* annual reduction in the poverty rate, then the poverty rate after one year is 0.50 · (1 – 0.10) = 0.450 = 45.0 percent, and the poverty rate after two years is 0.45 · (1 – 0.10) = 0.405 = 40.5 percent. On the other hand, if there is a *10.0-percentage-point* annual reduction in poverty, then the rate after one year is 0.50 – 0.10 = 0.40 = 40 percent, and the rate after two years is 0.40 – 0.10 = 0.30 = 30 percent.

[22] Of course, such a large annual reduction in poverty is unrealistic, but this is just an example to show how the scorecard can be used to estimate change.

Figure 11: Spreadsheet calculation of estimated annual net change in the head-count poverty rate and in the annual net number of poor people who rose above a poverty line with one sample scored twice

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ID of direct	Interview date		Years between	Number of household members			Member-years	Score		Poverty likelihood (%)		Estimated net change in
2	participant	Baseline	Follow-up	interviews	Baseline	Follow-up	Average	between	Baseline	Follow-up	Baseline	Follow-up	number of poor
3	1V0276FZ7	13-Jun-2020	13-Aug-2023	3.17 = (C3-B3)/365	9	8	8.50 = (E3+F3)/2	26.92 = D3*G3	17	22	94.8	86.5	−0.706 = G3*(L3-K3)/100
4	2W3120ZG8	30-Jun-2020	15-May-2023	2.87 = (C4-B4)/365	5	4	4.50 = (E4+F4)/2	12.93 = D4*G4	26	27	81.9	81.1	−0.036 = G4*(L4-K4)/100
5				**Average:**	7.0 = AVERAGE(E3:E4)	6.0 = AVERAGE(F3:F4)	**Sum:**	39.85 = SUM(H3:H4)					−0.742 = SUM(M3:M4)
6													
7						**Estimated net change in head-count poverty rate (percentage points), follow-up versus baseline:**								−5.7 = M5/(E5+F5)*100
8													
9								**Household-size-weighted average years between interviews:**					3.07 = H6/(E5+F5)
10													
11							**Estimated annual net change in head-count poverty rate (percentage points):**						−1.9 = M7/M9*100
12													
13									**Participating households at baseline:**				1,000
14									**Participating households at follow-up:**				700
15													
16								**Estimated average number of on-going participating people:**					5,600 = (E5*M13+F5*M14)/2
17													
18								**Estimated annual net change in the number of poor people:**					−104 = M16*M11/100
19	Rows of data are sorted by the ID of the direct participant.												

3.3.2 Annual net change in the number of poor people with one sample scored twice

For a pro-poor program, the bottom line is not the annual net change in the poverty rate but rather the annual net change in the number of poor participants who rise above a poverty line.

To calculate this, the first step is to estimate the average number of household members in the population of on-going households from baseline to follow-up, accounting for drop-out. In our example, the population of the in-coming households in the calendar-year 2020 cohort in 2020 was 1,000. By the end of the follow-up period of calendar-year 2023, 300 had dropped out, leaving 700. If drop-out took place at a constant pace and was unrelated to changes in poverty,[23] then an estimate of the average number of on-going participating people is the equal-weighted average of the number of participating people among households

---

[23] This assumption rarely holds. On the one hand, the households that benefit most from the program—and thus those for whom participation is most likely to cause a faster-than-otherwise decrease in poverty—may also be the least-likely to drop out, leading to too-high estimates of the reduction in poverty due to participation. On the other hand, households whose poverty decreases may be more likely to drop out if the benefits of continued participation fall as poverty decreases, leading to too-low estimates of impact. Unfortunately, there is no general way to adjust scorecard estimates to account for drop out that is related to changes in poverty. As in all decision-making, managers must use their experience and judgment to detect deviations from assumptions and then to account for them as best they can. This is true even though scorecard estimates are based on data and math. "Hard numbers" may not represent reality as accurately as they seem to, and only a manager's knowledge of context can detect and account for this. Managers should discount unreliable estimates when they have reasoned, explicit arguments to do so (Schreiner, 2016b). Of course, discretion also opens the door to abuse; faced with unexpectedly low estimates of poverty reduction, managers might quietly sweep them under the rug or blame them on a slow economy (even though they would not attribute high estimates of poverty reduction to a roaring economy). Ironically and sadly, such attempts to make a program look good by hiding or excusing undesired results destroys the results' value as feedback, harming the program's ability to fulfill its mission. If a program's funders fail to act like owners, then its employees—not its participants—often become its *de facto* beneficiaries (Schreiner, 1997).

interviewed at baseline and follow-up. In a given round, the number of participating people is the average household size for that round's interviewed households (in the example, 7.0 at baseline and 6.0 at follow-up), multiplied by the number of participating households in the population in the given round (1,000 at baseline and 700 at follow-up), divided by the number of survey rounds (two). This is $\frac{7.0 \cdot 1{,}000 + 6.0 \cdot 700}{1+1} = 5{,}600$ people.

The second and last step is to multiply the estimated annual change in the poverty rate (here, about –1.9 percentage points, or –0.019) by the estimated average number of on-going participants (here, 5,600). This gives an annual net change in the number of poor people by 150% of the national line of $-0.019 \cdot 5{,}600 \approx -104$ people.[24] This negative change is a reduction (improvement) in poverty; there are about 104 fewer poor people in participating households in this cohort each year.

3.3.3 Estimating a program's impact

Estimating *change* is not the same as an estimating a program's *impact*. It stands to reason that program participation is a force that does cause some change (be it an increase or decrease) in the poverty of its participants. At the same time, it is equally logical to expect that a large share of any change is caused by the many non-program forces that affect participants. On its own, the scorecard is like a bathroom scale; it can tell whether you lost weight in the past month, but not how much of the loss is due to eating right and exercising versus removing your coat and shoes.

This point is often forgotten, confused, or ignored, so it bears repeating: the scorecard estimates change, but it does not—on its own—identify the causes of change. In particular, estimating the impact of program participation requires knowledge or assumptions about what would have happened to participants if they had not been participants. This must come from beyond the scorecard.

What is a program manager to do? All decision-making hinges on forecasts of the expected impacts of possible choices, so a manager cannot pretend that merely estimating change is helpful without also inferring some impact. Yet there are diminishing returns to improving inferences of impact. At a minimum, a program should compare its estimated annual net change in the poverty rate of its on-going participants to third-party estimates for the country overall or for the area where the program works. A program can also look for signs that participants value (or

[24] This is a net figure; some start above the line and end below it, and vice versa.

expect to value) its services. Is the number of in-coming participants high or increasing? Is the drop-out rate low or decreasing? Are drop-outs mostly due to dissatisfaction or graduation? Is participation voluntary, without being a condition for some other linked benefit? Is the program the sole provider in its niche and area?

In short, decision-makers in pro-poor programs are called to do what good decision-makers must always do: weigh data and knowledge from a number of perspectives and sources—including scorecard estimates, but not *only* scorecard estimates—to inform reasoned guesses as to more or less what share of observed changes are due to program participation. Of course, the inevitable need for human wisdom/art may be disingenuously invoked as a cover for decision processes that do not take a program's pro-poor mission to heart. This is why the "scientific method"—that is, being transparent about inputs and reasoning so as to facilitate productive review and debate—makes sense even (or perhaps especially) for business decisions.[25]

3.3.4 Annual net change in poverty rates with two independent samples

Instead of interviewing the same sample of households at both baseline and follow-up, a program could draw a second, independent sample of households from the same population as that from which the baseline sample was drawn.[26] The head-count poverty rate for on-going participants in this new follow-up sample is estimated in the same way as for the baseline sample.

Continuing the example, suppose that a third household and a fourth household are sampled at follow-up. The third household is interviewed on March 3, 2023. It has three members, a score of 21, and a poverty likelihood by 150% of the national line of 89.8 percent (Figure 1).

The fourth household is interviewed on April 4, 2023. It has seven members, a score of 29, and a poverty likelihood of 78.1 percent.

As at baseline, the estimated head-count poverty rate at follow-up is the household-size-weighted average of the poverty likelihoods of the sampled households: $\frac{3 \cdot 0.898 + 7 \cdot 0.781}{3+7} \approx \frac{2.69 + 5.47}{10} \approx 0.816 = 81.6$ percent.

[25] Schreiner (2016b) and Schreiner (2014).

[26] By chance, some households may end up in both samples.

The estimated annual net change in the head-count poverty rate of on-going participants is then the difference between the poverty-rate estimates at follow-up (81.6 percent) versus at baseline (90.2 percent),[27] divided by the difference (in years) between the household-size-weighted average of follow-up interview dates (March 25, 2023) versus the household-size-weighted average of baseline interview dates (June 10, 2020). These two average dates differ by about 1,009 days or 2.77 years.

The estimated annual net change in the head-count poverty rate is the difference between the poverty-rate estimates at follow-up versus baseline, divided by the difference in the average years between interviews in the two rounds. For 150% percent of the national line, this is (81.6 – 90.2) ÷ 2.77 ≈ –3.1 percentage points per year.

In practice, the calculations are done in an app or a spreadsheet (Figure 12).

[27] With two independent samples, the estimation error in each of the two snapshot estimates washes out, so it is not explicitly included in the calculation.

Figure 12: Spreadsheet calculation of estimated annual net change in a head-count poverty rate and in the annual net number of poor people who rise above a poverty line with two independent samples

	A	B	C	D	E	F	G	H
1	Survey	ID of direct participant	Interview date	Number of household members	Interview date x Number of household members	Score	Poverty likelihood (%)	Estimated number of poor household members
2	Baseline	1V0276FZ7	13-Jun-2020	9	31-Jan-2984 = C2*D2	17	94.8	8.53 = D2*G2/100
3	Baseline	2W3120ZG8	30-Jun-2020	5	02-Jul-2502 = C2*D2	26	81.9	4.10 = D3*G3/100
4	Follow-up	3XA76T21L	3-Mar-2023	3	07-Jul-2269 = C2*D2	21	89.8	2.69 = D4*G4/100
5	Follow-up	4Y8Y3EQS9	4-Apr-2023	7	27-Oct-2762 = C2*D2	29	78.1	5.47 = D5*G5/100
6		**Sum baseline:**		14 = SUM(D2:D3)				12.63 = SUM(H2:H3)
7		**Sum follow-up:**		10 = SUM(D4:D5)				8.16 = SUM(H4:H5)
8		**Average baseline:**		7.0 = AVERAGE(D2:D3)	19-Jun-2020 = SUM(E2:E3)/D6			
9		**Average follow-up:**		5.0 = AVERAGE(D4:D5)	25-Mar-2023 = SUM(E4:E5)/D7			
10								
11					**Estimated baseline poverty rate (%):**			90.2 = H6/D6*100
12					**Estimated follow-up poverty rate (%):**			81.6 = H7/D7*100
13								
14				**Average years between follow-up and baseline interviews:**				2.77 = (E9-E8)/365
15								
16			**Estimated annual net change in head-count poverty rate (percentage points):**					–3.1 = (H12-H11)/H14
17								
18					**Participating households at baseline:**			1,000
19					**Participating households at follow-up:**			700
20								
21				**Estimated average number of on-going participating people:**				5,250 = (D8*H18+D9*H19)/2
22								
23				**Estimated annual net change in the number of poor people:**				–163 = H21*H16/100
24	Rows of data are sorted by Round, then by Interview date, then by Direct participant ID.							

3.3.5 Annual net change in the number of poor people with two independent samples

For a pro-poor program, the bottom line is not the annual net change in the poverty rate but rather the annual net change in the number of poor participants who rise above a poverty line.

To calculate this, the first step is to estimate the average number of household members in the population of on-going households from baseline to follow-up, accounting for drop-out. In our example, the population of the baseline 2020 cohort in 2020 is 1,000 in-coming households. By the end of the follow-up period 2023, 300 households have dropped out, leaving 700. If drop-out took place at a constant pace and was unrelated to changes in poverty, then an estimate of the average number of on-going participating people is the equal-weighted average of the number of participating people among households interviewed at baseline and follow-up. In a given round, the number of participating people is the average household size for that round's interviewed households (in our example, 7.0 at baseline and 5.0 at follow-up), multiplied by the number of participating households in the population in the given round (1,000 at baseline and 700 at follow-up), and divided by two (the number of rounds). This is $\frac{7.0 \cdot 1{,}000 + 5.0 \cdot 700}{1+1} = 5{,}250$ people.

The second and last step is to multiply the estimated annual net change in the head-count poverty rate (here, –3.1 percentage points, or –0.031) by the estimated number of on-going participants (here, 5,250). For 150% of the national line, this gives an annual net change in the number of poor people of $-0.031 \cdot 5{,}250 \approx -164$ people per year. This negative change is a (non-compounded) reduction in poverty; the number of poor people in participating households decreases (improves) by 164 each year.

Given the scorecard's assumptions, both approaches to estimating change over time—one sample scored twice, and two independent samples—are unbiased. In general, the two approaches give different estimates (as in this example) because they interview different households at different times. All else constant, scoring one sample twice has smaller margins of error, but there may be context-specific reasons (related to operational costs or non-sampling errors) to score two independent samples.

4. How to design scorecard surveys and samples

To design a scorecard survey and its sample, a program must decide:[28]

- Who will do interviews
- Where and how to do interviews
- How to record responses and scores
- How to calculate estimates and report/analyze them
- Which participating households to interview
- How many participating households to interview
- How frequently to do surveys
- Whether to track a population across multiple time periods
- Whether to interview the same participants twice

Decisions should follow from the program's goals, the business decisions to be informed, and the budget. The central goals of the design are to:

- Inform issues that matter to the program
- Make sure that the sample is representative of a well-defined population

4.1 Who will do interviews

The enumerators who interview participating households must be trained to follow the "Interview Guide". Enumerators may be:

- Program employees
- Contractors

4.2 Where and how to do interviews

Interviews should be:

- In-person, and
- At the sampled household's residence, and
- With an enumerator trained to follow the "Interview Guide"

This is the only recommended way. It follows Tanzania's NBS in the 2018 HBS, so it provides the most-accurate and most-consistent data (and thus the best estimates).

[28] IRIS Center (2007) and Toohig (2008) also discuss this topic, covering sampling, budgeting, training, logistics, interviewing, piloting, and recording data.

Of course, it is possible to do interviews in non-recommended ways such as:

- Without an enumerator (such as by respondents' filling out paper or web forms on their own or responding to questions sent via e-mail, texts, or robo-calls)
- Away from home (such as a program's service point or a local meeting place)
- Not in-person (such as with an enumerator by phone)

While non-recommended methods may reduce costs, they also affect responses[29] and thus reduce the accuracy of estimates. This is why interviewing by a trained enumerator at the residence is recommended.

In some contexts—such as when a program's field agents do not already visit participants at their residences anyway as part of their normal work—a program might be willing to trade accuracy for a lower-cost, non-recommended approach. The business wisdom of this depends on context-specific factors that each program must judge for itself. To judge carefully, a program that is considering a non-recommended method should do a small test to see how responses differ when compared with a trained enumerator at the residence. Furthermore, all reporting should discuss the possible consequences of the non-recommended method.

4.3 How to record responses and scores

Responses and scores may be recorded by enumerators on:

- Paper, and then keyed into a database or spreadsheet at an office
- Mobile devices, and then uploaded to a database[30]

4.4 How to calculate estimates and report and analyze them

Analysts can calculate estimates by plugging data into spreadsheets (following the examples in Section 3) or with the spreadsheet-based PovIt!™-brand reporting app. Schreiner (2014) describes how to report and analyze scorecard estimates.

[29] Schreiner, 2015.

[30] Scorocs can help set up a system to collect data with mobile devices or to transfer data from paper forms into a database at the office. Support is also available for calculating estimates and for reporting and analysis.

4.5 Which participating households to interview

Given a population relevant for a particular business decision, the participating households to be interviewed can be:

- All relevant participants (a census)
- A representative sample of relevant participants
- All relevant participants in a representative sample of relevant service points and/or in a representative sample of relevant field agents
- A representative sample of relevant participants in a representative sample of relevant service points and/or in a representative sample of relevant field agents

A census is rarely necessary, except for very small programs. Nevertheless, it may be less costly to interview all in-coming households as a standard part of in-take rather than managing who gets scored and who does not.

4.6 How many participating households to interview

If not determined by other factors, the number of participating households to interview can be derived from sample-size formulas to achieve a desired confidence level for a desired margin of error (Annex 6).

The focus of sample design, however, should be less on having enough interviews to achieve some arbitrary level of statistical significance and more on having a representative sample from a well-defined population that is relevant for informing decisions that matter to the program.

In practice, non-sampling errors in implementation and in the definition of the population often matter at least as much as errors due to smaller samples. Programs are often concerned about sample size, but as there is no point in deriving the ideal sample size unless proportional effort goes to mitigating other sources of error and then accounting for margins of error in the analysis stage. Of course, smaller samples produce less-reliable estimates. In practice, however, almost no one reports or considers margins of error (even though they should), and estimates based on at least 1,000 interviews will rarely raise eyebrows (Annex 6).

4.7 How frequently to do surveys

The frequency of scorecard surveys can be:

- As a once-off project (precluding estimating change)
- Every three years (or at any other fixed or variable time interval, allowing estimating change)
- Each time a field agent visits a participant at home (allowing estimating change)

4.8 Whether to track a population across periods

The scorecard can estimate changes in poverty across periods, but not all programs want to do this. Many programs want to assess poverty only for in-coming participants.

4.9 Whether to interview the same participants twice

If a scorecard is applied more than once in order to estimate changes in poverty, then it can be applied with:

- One sample of participants, all of whom are scored at both baseline and follow-up
- Two samples of participants from the same population, with the first sample scored at baseline and the second sample scored at follow-up.

All else constant, scoring one sample twice gives estimates with smaller margins of error. It may also be less costly at follow-up, given that the sampled households have already been tracked down at home at baseline. Also, the follow-up round could be based on a random sample of the households interviewed at baseline.

4.10 Example of survey design and implementation in Bangladesh

An example set of choices is illustrated by the microfinance arms of BRAC and ASA, two pro-poor titans in Bangladesh who each have about 7 million participating households and who made plans to apply the scorecard for Bangladesh[31] with a sample of about 25,000 participants each.

Their design is that all loan officers in a random sample of branches score all participants each time the loan officers visit a homestead (about once a year) as part of their standard due diligence prior to loan disbursement. The loan officers record responses on paper in the field before sending the forms to a central office to be entered into a database and converted to poverty likelihoods for further analysis.

[31] Schreiner, 2013b.

5. How to use scores for targeting

When a program uses the scorecard for segmenting (*targeting*) participants for differentiated treatment based on poverty, people in households with scores at or below a cut-off are labeled *targeted* and given one type of treatment. People in households with scores above a cut-off are labeled *non-targeted* and given another type of treatment.[32]

Households that score at or below a given cut-off should be labeled as *targeted*,[33] not as *poor*.[34]

Targeting is successful to the extent to which people truly below a poverty line are targeted (*inclusion*) or people truly above a poverty line are not targeted (*exclusion*). Of course, no poverty-assessment tool is perfect, and targeting is unsuccessful to the extent to which people truly below a poverty line are not targeted (*undercoverage*) or people truly above a poverty line are targeted (*leakage*).

---

[32] *Targeting status* (having a score at or below a targeting cut-off) is not the same concept as *poverty status* (having consumption expenditure below a poverty line). Poverty status is a fact that is defined by whether consumption expenditure is below a poverty line as directly measured by a survey. In contrast, targeting status is a program's policy choice that depends on a cut-off and on an indirect estimate from a scorecard.

[33] Other labels can be meaningful as long as they describe the segment and do not confuse targeting status (having a score below a program-selected cut-off) with poverty status (having consumption expenditure below an externally-defined poverty line). Examples include: *Groups A, B, and C*; *People with scores of 29 or less, 30 to 69, or 70 or more*; and *People who qualify for reduced fees, or who do not qualify*.

[34] After all, unless all targeted households have poverty likelihoods of 100 percent, it is likely that some of them are non-poor (their consumption expenditure is above a given poverty line). In the context of the scorecard, the terms *poor* and *non-poor* have specific, consumption-based definitions. Using these same terms for targeting status is incorrect and misleading.

Figure 13 below depicts these four possible targeting outcomes. Targeting accuracy varies by the cut-off score. A higher cut-off has better inclusion and better undercoverage (but worse exclusion and worse leakage). In contrast, a lower cut-off has worse inclusion and worse undercoverage (but better exclusion and better leakage).

Figure 13: Possible targeting outcomes

Observed poverty status	Targeting segment: Targeted	Targeting segment: Non-targeted
Poor	**Inclusion** Poor correctly targeted	**Undercoverage** Poor mistakenly not targeted
Non-poor	**Leakage** Non-poor mistakenly targeted	**Exclusion** Non-poor correctly not targeted

Programs should weigh these trade-offs when setting a cut-off. A formal way to do this is to assign net benefits—based on a program's values and mission—to each of the four possible targeting outcomes and then to choose the cut-off that maximizes the sum of net benefits.[35]

The five tables below show the scorecard's targeting outcomes by poverty line and by score cut-off for people in Tanzania:

- Figure 14: Inclusion (% people who are poor and correctly targeted)
- Figure 15: Undercoverage (% people who are poor but mistakenly not targeted)
- Figure 16: Leakage (% people who are not poor but mistakenly targeted)
- Figure 17: Exclusion (% people who are not poor and correctly not targeted)
- Figure 18: Hit rate (% people correctly targeted, that is, inclusion plus exclusion)

For a given score cut-off, each of the five figures below also show the share of all people who are targeted.

[35] Adams and Hand, 2000; Hoadley and Oliver, 1998.

Figure 14: Inclusion (% people who are poor and correctly targeted)

Targeting cut-off	% all people who are targeted	Inclusion (%)														
		National				Intl. 2011 PPP				Percentile-based lines						
		Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
<=17	5.1	1.7	3.8	4.8	4.9	4.8	5.1	5.1	5.1	2.6	3.3	4.5	4.8	4.9	5.1	5.1
<=21	9.9	2.6	6.5	9.3	9.7	9.1	9.9	9.9	9.9	3.7	5.7	8.2	9.1	9.5	9.9	9.9
<=24	14.9	3.7	9.4	13.6	14.4	13.2	14.7	14.9	14.9	4.9	7.8	12.1	13.4	14.1	14.7	14.9
<=26	19.5	4.6	11.5	17.5	18.8	17.1	19.2	19.5	19.5	5.9	9.6	15.3	17.3	18.2	19.2	19.5
<=28	24.7	5.1	13.8	22.0	23.7	21.1	24.2	24.6	24.6	6.6	11.2	18.8	21.5	22.8	24.3	24.5
<=30	29.2	5.8	15.5	25.3	27.7	24.3	28.4	29.1	29.2	7.3	12.5	21.6	24.8	26.4	28.5	29.1
<=32	33.8	6.2	17.4	29.1	32.2	27.7	33.0	33.7	33.8	7.9	14.2	24.5	28.2	30.4	33.1	33.7
<=34	39.5	6.6	19.0	33.2	37.0	31.4	38.3	39.4	39.5	8.6	15.3	27.2	31.9	34.7	38.5	39.4
<=36	43.8	7.2	20.4	36.3	40.9	34.1	42.3	43.7	43.8	9.2	16.4	29.7	34.7	38.2	42.6	43.6
<=38	48.4	7.5	21.5	39.3	44.8	37.1	46.4	48.1	48.4	9.5	17.2	32.0	37.7	41.6	46.7	48.0
<=40	53.6	7.8	22.8	42.0	48.6	39.6	50.8	53.1	53.6	9.9	18.3	34.0	40.2	44.9	51.2	52.9
<=42	58.0	7.9	23.9	44.2	52.0	41.7	54.6	57.5	58.0	10.1	18.7	35.6	42.3	47.6	55.2	57.3
<=45	63.9	8.0	24.8	47.3	56.3	44.3	59.5	63.0	63.8	10.2	19.2	37.3	44.9	51.0	60.0	62.8
<=48	68.4	8.0	25.2	49.0	59.2	45.5	62.6	67.4	68.3	10.2	19.4	38.1	46.3	53.1	63.7	67.1
<=51	74.0	8.0	25.7	50.8	62.5	47.0	66.2	72.9	74.0	10.2	19.7	39.1	47.8	55.4	67.9	72.5
<=54	78.9	8.1	26.2	51.8	64.9	47.9	69.1	77.2	78.8	10.2	19.8	39.8	48.7	56.8	71.0	76.7
<=57	83.6	8.1	26.4	52.9	67.2	48.6	71.6	81.4	83.6	10.3	20.0	40.2	49.4	58.1	73.7	80.6
<=61	88.2	8.1	26.5	53.5	69.0	49.1	73.7	85.2	88.2	10.3	20.1	40.5	50.0	59.1	76.2	84.1
<=66	93.9	8.1	26.8	54.2	70.5	49.6	75.5	89.4	93.7	10.3	20.1	40.9	50.5	59.8	78.4	88.0
<=100	100.0	8.1	26.8	54.4	71.4	49.6	76.8	92.8	99.7	10.3	20.1	40.9	50.6	60.1	79.8	91.1

Scorecard applied to the validation sample.

Figure 15: Undercoverage (% people who are poor but mistakenly not targeted)

Targeting cut-off	% all people who are targeted	Undercoverage (%)														
		National				Intl. 2011 PPP				Percentile-based lines						
		Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
<=17	5.1	6.5	23.0	49.6	66.5	44.8	71.7	87.7	94.6	7.7	16.8	36.4	45.8	55.1	74.8	86.0
<=21	9.9	5.5	20.3	45.1	61.7	40.6	66.9	82.8	89.8	6.6	14.5	32.7	41.5	50.5	69.9	81.1
<=24	14.9	4.4	17.5	40.8	57.0	36.4	62.1	77.9	84.9	5.4	12.3	28.9	37.2	46.0	65.1	76.2
<=26	19.5	3.5	15.3	36.9	52.6	32.6	57.6	73.3	80.3	4.4	10.5	25.6	33.3	41.9	60.6	71.6
<=28	24.7	3.0	13.0	32.4	47.7	28.5	52.6	68.2	75.1	3.7	8.9	22.2	29.1	37.3	55.6	66.5
<=30	29.2	2.4	11.3	29.1	43.7	25.3	48.4	63.7	70.6	3.0	7.6	19.3	25.8	33.6	51.3	62.0
<=32	33.8	1.9	9.4	25.3	39.2	21.9	43.8	59.0	65.9	2.4	6.0	16.4	22.4	29.6	46.7	57.4
<=34	39.5	1.5	7.8	21.2	34.4	18.3	38.5	53.3	60.2	1.7	4.8	13.7	18.7	25.4	41.4	51.7
<=36	43.8	0.9	6.4	18.1	30.5	15.5	34.4	49.1	55.9	1.1	3.7	11.3	15.9	21.9	37.2	47.5
<=38	48.4	0.6	5.3	15.1	26.7	12.5	30.4	44.7	51.3	0.8	2.9	9.0	12.9	18.4	33.1	43.1
<=40	53.6	0.3	4.0	12.4	22.8	10.0	26.0	39.6	46.1	0.4	1.8	7.0	10.4	15.1	28.7	38.1
<=42	58.0	0.2	2.9	10.2	19.4	8.0	22.1	35.3	41.7	0.3	1.4	5.3	8.3	12.5	24.7	33.8
<=45	63.9	0.1	2.0	7.0	15.1	5.4	17.3	29.7	35.9	0.1	0.9	3.6	5.7	9.0	19.8	28.2
<=48	68.4	0.1	1.6	5.4	12.3	4.2	14.2	25.4	31.4	0.1	0.7	2.9	4.3	6.9	16.1	24.0
<=51	74.0	0.1	1.1	3.6	8.9	2.7	10.5	19.9	25.7	0.1	0.5	1.8	2.8	4.7	11.9	18.6
<=54	78.9	0.0	0.7	2.6	6.5	1.8	7.7	15.5	20.9	0.1	0.3	1.2	1.9	3.2	8.9	14.4
<=57	83.6	0.0	0.4	1.5	4.2	1.1	5.2	11.4	16.2	0.0	0.2	0.7	1.2	1.9	6.1	10.5
<=61	88.2	0.0	0.3	0.8	2.4	0.6	3.1	7.6	11.6	0.0	0.1	0.4	0.6	1.0	3.7	7.0
<=66	93.9	0.0	0.0	0.2	0.9	0.1	1.3	3.4	6.0	0.0	0.0	0.0	0.1	0.2	1.4	3.1
<=100	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Scorecard applied to the validation sample.

Figure 16: Leakage (% people who are not poor but mistakenly targeted)

Targeting cut-off	% all people who are targeted	Leakage (%)														
		National				Intl. 2011 PPP				Percentile-based lines						
		Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
<=17	5.1	3.4	1.3	0.3	0.2	0.3	0.0	0.0	0.0	2.5	1.8	0.6	0.3	0.2	0.0	0.0
<=21	9.9	7.3	3.4	0.7	0.2	0.9	0.1	0.0	0.0	6.2	4.3	1.7	0.8	0.4	0.0	0.0
<=24	14.9	11.1	5.5	1.3	0.4	1.6	0.2	0.0	0.0	9.9	7.1	2.8	1.5	0.8	0.2	0.0
<=26	19.5	14.9	8.0	2.0	0.7	2.4	0.3	0.0	0.0	13.6	10.0	4.2	2.3	1.4	0.3	0.0
<=28	24.7	19.5	10.9	2.7	0.9	3.5	0.5	0.1	0.0	18.1	13.3	5.9	3.2	1.9	0.4	0.1
<=30	29.2	23.4	13.7	3.9	1.4	4.9	0.8	0.1	0.0	21.9	16.5	7.6	4.4	2.8	0.7	0.1
<=32	33.8	27.7	16.5	4.8	1.7	6.1	0.9	0.1	0.0	26.0	19.6	9.4	5.6	3.4	0.7	0.1
<=34	39.5	32.9	20.5	6.4	2.6	8.1	1.3	0.1	0.0	30.9	24.1	12.3	7.6	4.8	1.1	0.1
<=36	43.8	36.6	23.4	7.5	2.9	9.7	1.5	0.1	0.0	34.6	27.3	14.2	9.2	5.6	1.2	0.2
<=38	48.4	40.9	26.9	9.2	3.7	11.3	2.0	0.3	0.0	38.9	31.1	16.5	10.8	6.8	1.7	0.4
<=40	53.6	45.8	30.8	11.6	5.0	14.0	2.9	0.5	0.0	43.7	35.2	19.6	13.4	8.7	2.4	0.7
<=42	58.0	50.1	34.1	13.8	6.1	16.4	3.4	0.6	0.0	48.0	39.3	22.4	15.8	10.5	2.9	0.8
<=45	63.9	55.8	39.0	16.5	7.6	19.6	4.4	0.8	0.0	53.7	44.6	26.5	18.9	12.8	3.8	1.0
<=48	68.4	60.3	43.2	19.4	9.2	22.9	5.8	0.9	0.0	58.1	48.9	30.3	22.1	15.2	4.7	1.3
<=51	74.0	66.0	48.3	23.3	11.5	27.1	7.8	1.2	0.0	63.8	54.3	34.9	26.2	18.6	6.1	1.6
<=54	78.9	70.8	52.7	27.0	14.0	31.0	9.8	1.6	0.0	68.6	59.0	39.1	30.2	22.1	7.9	2.2
<=57	83.6	75.5	57.2	30.8	16.5	35.1	12.0	2.2	0.0	73.3	63.6	43.4	34.2	25.5	9.9	3.1
<=61	88.2	80.1	61.7	34.7	19.2	39.1	14.6	3.0	0.1	77.9	68.1	47.7	38.2	29.1	12.1	4.2
<=66	93.9	85.7	67.1	39.7	23.3	44.3	18.3	4.5	0.1	83.5	73.7	53.0	43.3	34.0	15.4	5.9
<=100	100.0	91.9	73.2	45.6	28.6	50.4	23.2	7.2	0.3	89.7	79.9	59.1	49.4	39.9	20.2	8.9

Scorecard applied to the validation sample.

Figure 17: Exclusion (% people who are not poor and correctly not targeted)

Targeting cut-off	% all people who are targeted	Exclusion (%)														
		National				Intl. 2011 PPP				Percentile-based lines						
		Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
<=17	5.1	88.4	71.9	45.4	28.4	50.1	23.2	7.2	0.3	87.2	78.1	58.5	49.1	39.8	20.2	8.9
<=21	9.9	84.6	69.8	44.9	28.3	49.5	23.2	7.2	0.3	83.5	75.6	57.3	48.6	39.5	20.1	8.9
<=24	14.9	80.7	67.7	44.3	28.1	48.7	23.0	7.2	0.3	79.7	72.8	56.2	47.9	39.1	20.0	8.9
<=26	19.5	77.0	65.2	43.6	27.9	47.9	22.9	7.2	0.2	76.1	69.9	54.9	47.1	38.6	19.9	8.9
<=28	24.7	72.4	62.3	42.9	27.7	46.8	22.7	7.1	0.2	71.6	66.6	53.2	46.2	38.1	19.8	8.8
<=30	29.2	68.4	59.5	41.7	27.1	45.5	22.4	7.1	0.2	67.8	63.3	51.5	44.9	37.2	19.5	8.8
<=32	33.8	64.2	56.7	40.8	26.9	44.3	22.3	7.1	0.2	63.7	60.3	49.7	43.7	36.5	19.4	8.8
<=34	39.5	59.0	52.7	39.2	26.0	42.2	21.9	7.1	0.2	58.7	55.8	46.8	41.8	35.1	19.1	8.8
<=36	43.8	55.3	49.8	38.1	25.7	40.6	21.7	7.1	0.2	55.0	52.6	44.9	40.2	34.3	18.9	8.7
<=38	48.4	50.9	46.3	36.4	24.9	39.0	21.2	6.9	0.2	50.7	48.8	42.6	38.6	33.1	18.5	8.5
<=40	53.6	46.1	42.4	34.0	23.6	36.4	20.4	6.7	0.2	46.0	44.7	39.5	36.0	31.3	17.7	8.3
<=42	58.0	41.8	39.0	31.8	22.5	34.0	19.8	6.6	0.2	41.7	40.6	36.7	33.6	29.5	17.3	8.2
<=45	63.9	36.0	34.2	29.1	21.0	30.8	18.8	6.4	0.2	36.0	35.3	32.6	30.5	27.1	16.3	7.9
<=48	68.4	31.5	30.0	26.2	19.4	27.5	17.5	6.3	0.2	31.5	31.0	28.8	27.3	24.7	15.5	7.7
<=51	74.0	25.9	24.9	22.3	17.0	23.3	15.4	6.1	0.2	25.9	25.6	24.1	23.2	21.3	14.0	7.4
<=54	78.9	21.1	20.5	18.6	14.6	19.4	13.4	5.6	0.2	21.0	20.8	20.0	19.2	17.9	12.3	6.7
<=57	83.6	16.3	15.9	14.8	12.1	15.3	11.2	5.0	0.2	16.3	16.2	15.7	15.2	14.4	10.3	5.9
<=61	88.2	11.8	11.5	10.9	9.4	11.2	8.6	4.2	0.2	11.8	11.8	11.3	11.2	10.8	8.1	4.8
<=66	93.9	6.1	6.1	5.9	5.2	6.1	4.9	2.8	0.1	6.1	6.1	6.1	6.1	5.9	4.7	3.1
<=100	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Scorecard applied to the validation sample.

Figure 18: Hit rate (% people correctly targeted, that is, inclusion plus exclusion)

Targeting cut-off	% all people who are targeted	Hit rate (= Inclusion + Exclusion) (%)														
		National				Intl. 2011 PPP				Percentile-based lines						
		Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
<=17	5.1	90.1	75.7	50.2	33.3	54.9	28.3	12.3	5.4	89.8	81.5	63.0	54.0	44.7	25.2	14.0
<=21	9.9	87.2	76.3	54.2	38.0	58.5	33.0	17.2	10.2	87.2	81.3	65.6	57.7	49.0	30.0	18.8
<=24	14.9	84.5	77.0	57.9	42.6	62.0	37.7	22.1	15.1	84.7	80.6	68.3	61.3	53.2	34.7	23.8
<=26	19.5	81.6	76.7	61.2	46.7	65.0	42.2	26.7	19.7	82.0	79.5	70.2	64.4	56.7	39.1	28.4
<=28	24.7	77.5	76.1	64.9	51.4	67.9	46.9	31.7	24.9	78.2	77.7	71.9	67.7	60.8	44.0	33.3
<=30	29.2	74.2	75.0	67.1	54.9	69.8	50.8	36.2	29.4	75.1	75.9	73.1	69.7	63.6	48.0	37.9
<=32	33.8	70.4	74.1	69.9	59.1	72.0	55.3	40.9	34.1	71.6	74.5	74.2	72.0	67.0	52.5	42.5
<=34	39.5	65.6	71.8	72.4	63.0	73.6	60.2	46.6	39.8	67.3	71.2	74.0	73.7	69.8	57.6	48.2
<=36	43.8	62.5	70.2	74.4	66.6	74.8	64.0	50.8	44.1	64.2	69.0	74.5	74.9	72.5	61.5	52.3
<=38	48.4	58.4	67.8	75.7	69.7	76.2	67.6	55.0	48.7	60.2	66.0	74.5	76.3	74.7	65.2	56.5
<=40	53.6	54.0	65.2	76.1	72.2	76.0	71.1	59.9	53.8	55.9	63.0	73.4	76.2	76.2	68.9	61.2
<=42	58.0	49.7	62.9	76.0	74.5	75.7	74.5	64.1	58.3	51.8	59.3	72.3	75.9	77.0	72.5	65.4
<=45	63.9	44.1	59.0	76.4	77.3	75.0	78.3	69.5	64.0	46.2	54.4	69.9	75.4	78.1	76.4	70.7
<=48	68.4	39.6	55.2	75.2	78.5	73.0	80.1	73.7	68.5	41.8	50.4	66.9	73.6	77.9	79.2	74.8
<=51	74.0	33.9	50.6	73.1	79.5	70.3	81.7	78.9	74.2	36.1	45.2	63.3	71.0	76.7	82.0	79.8
<=54	78.9	29.2	46.6	70.4	79.5	67.2	82.4	82.8	79.1	31.3	40.6	59.7	67.9	74.7	83.2	83.4
<=57	83.6	24.4	42.3	67.7	79.3	63.8	82.8	86.4	83.8	26.6	36.2	55.9	64.6	72.6	84.0	86.4
<=61	88.2	19.9	38.0	64.5	78.4	60.3	82.3	89.4	88.4	22.1	31.8	51.9	61.2	69.9	84.3	88.8
<=66	93.9	14.3	32.9	60.1	75.8	55.6	80.4	92.2	93.8	16.5	26.3	47.0	56.6	65.7	83.1	91.1
<=100	100.0	8.1	26.8	54.4	71.4	49.6	76.8	92.8	99.7	10.3	20.1	40.9	50.6	60.1	79.8	91.1

Scorecard applied to the validation sample.

For an example cut-off of 32 or less in to the previous figures, 33.8 percent of all people are targeted, and outcomes for 150% of the national line in the validation sample are:

- Inclusion: 29.1 percent are below the line and correctly targeted
- Undercoverage: 25.3 percent are below the line and mistakenly not targeted
- Leakage: 4.8 percent are above the line and mistakenly targeted
- Exclusion: 40.8 percent are above the line and correctly not targeted

Increasing the cut-off to 34 or less increases the share of of all people targeted to 39.5 percent. The higher cut-off improves inclusion and undercoverage but worsens leakage and exclusion:

- Inclusion: 33.2 percent are below the line and correctly targeted
- Undercoverage: 21.2 percent are below the line and mistakenly not targeted
- Leakage: 6.4 percent are above the line and mistakenly targeted
- Exclusion: 39.2 percent are above the line and correctly not targeted

Which cut-off is preferred depends on the sum of net benefits. If each targeting outcome has a per-person benefit or cost, then total net benefit for a given cut-off is:

Benefit per person correctly included	x	People correctly included	–
Cost per person mistakenly not covered	x	People mistakenly not covered	–
Cost per person mistakenly leaked	x	People mistakenly leaked	+
Benefit per person correctly excluded	x	People correctly excluded.	

To set an optimal cut-off, a program would:

- Assign benefits and costs to possible outcomes, based on its values and mission
- Tally total net benefits for each cut-off using the figures above for a chosen poverty line
- Select the cut-off with the highest total net benefit

The most difficult step is assigning benefits and costs to targeting outcomes. A pro-poor program that uses targeting—with or without the scorecard—should thoughtfully consider how it values successful inclusion and exclusion versus errors of undercoverage and leakage. It is healthy to go through a process of thinking explicitly and intentionally about how targeting outcomes are valued.

A common choice of benefits and costs is the *hit rate*, where total net benefit is the number of people correctly included or correctly excluded:

Hit rate =	1	x	People correctly included	–
	0	x	People mistakenly undercovered	–
	0	x	People mistakenly leaked	+
	1	x	People correctly excluded.	

Figure 18 shows the scorecard's hit rate for all cut-offs and poverty lines. For the example of 150% of the national line in the validation sample, total net benefit under the hit rate for a cut-off of 32 or less is 69.9 percent. Among the 33.8 percent of all Tanzanians who are targeted, about seven in ten are correctly classified.

The hit rate weighs the successful inclusion of people below a poverty line the same as the successful exclusion of people above the line. If a program values inclusion more (say, twice as much) than exclusion, then it can reflect this by setting the benefit for inclusion to 2 and the benefit for exclusion to 1. Then the chosen cut-off will maximize (2 x people correctly included) + (1 x people correctly excluded).

As an alternative to assigning benefits and costs to targeting outcomes and then setting a score cut-off to maximize net benefits, a pro-poor program could set cut-offs based on aspects of targeting accuracy from the three figures below:

- Figure 19: Share of targeted people who are poor
- Figure 20: Poor people correctly targeted per non-poor person mistakenly targeted
- Figure 21: Share of poor people who are targeted

Figure 19: Share of targeted people who are poor

Targeting cut-off	% all people who are targeted	% targeted people who are poor														
		National				Intl. 2011 PPP				Percentile-based lines						
		Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
<=17	5.1	32.7	75.0	95.0	96.8	94.9	100.0	100.0	100.0	50.8	65.4	88.6	94.9	96.8	100.0	100.0
<=21	9.9	26.6	65.8	93.3	97.5	91.0	99.5	100.0	100.0	37.6	56.9	82.7	91.8	95.7	99.7	99.9
<=24	14.9	25.0	62.9	91.2	97.1	89.1	98.8	100.0	100.0	33.1	52.3	81.1	89.9	94.5	99.0	99.9
<=26	19.5	23.8	59.0	89.9	96.4	87.5	98.6	99.9	99.9	30.4	49.1	78.6	88.5	93.1	98.7	99.8
<=28	24.7	20.9	55.9	89.1	96.2	85.6	98.0	99.6	99.9	26.8	45.6	76.1	87.1	92.3	98.4	99.5
<=30	29.2	19.7	53.2	86.8	95.1	83.3	97.3	99.7	99.9	25.0	43.1	74.0	84.8	90.5	97.7	99.6
<=32	33.8	18.3	51.4	85.9	95.1	81.9	97.4	99.7	99.9	23.3	42.0	72.4	83.3	89.9	97.8	99.6
<=34	39.5	16.8	48.2	83.9	93.5	79.4	96.8	99.8	99.9	21.8	38.9	68.9	80.7	87.7	97.3	99.7
<=36	43.8	16.5	46.6	82.9	93.3	77.8	96.6	99.7	100.0	21.0	37.6	67.6	79.1	87.1	97.2	99.5
<=38	48.4	15.5	44.4	81.1	92.4	76.6	95.8	99.4	100.0	19.6	35.6	65.9	77.8	85.9	96.5	99.1
<=40	53.6	14.6	42.5	78.4	90.7	73.9	94.7	99.1	100.0	18.5	34.2	63.4	75.0	83.8	95.5	98.8
<=42	58.0	13.7	41.2	76.2	89.6	71.8	94.1	99.0	100.0	17.3	32.2	61.5	72.9	81.9	95.1	98.7
<=45	63.9	12.6	38.9	74.1	88.2	69.3	93.1	98.7	99.9	16.0	30.1	58.5	70.4	79.9	94.0	98.4
<=48	68.4	11.7	36.8	71.6	86.5	66.5	91.6	98.6	99.9	15.0	28.4	55.7	67.7	77.7	93.2	98.2
<=51	74.0	10.9	34.8	68.6	84.4	63.5	89.5	98.4	99.9	13.8	26.6	52.8	64.6	74.8	91.7	97.9
<=54	78.9	10.3	33.2	65.7	82.3	60.7	87.5	97.9	99.9	13.0	25.1	50.4	61.7	72.0	90.0	97.2
<=57	83.6	9.7	31.5	63.2	80.3	58.1	85.6	97.3	99.9	12.3	23.9	48.1	59.1	69.5	88.2	96.3
<=61	88.2	9.2	30.1	60.7	78.2	55.6	83.5	96.6	99.9	11.7	22.8	45.9	56.7	67.0	86.3	95.3
<=66	93.9	8.7	28.5	57.7	75.1	52.8	80.5	95.2	99.8	11.0	21.4	43.6	53.8	63.7	83.5	93.8
<=100	100.0	8.1	26.8	54.4	71.4	49.6	76.8	92.8	99.7	10.3	20.1	40.9	50.6	60.1	79.8	91.1

Scorecard applied to the validation sample.

Figure 20: Poor people correctly targeted per non-poor person mistakenly targeted

Targeting cut-off	% all people who are targeted	Poor people targeted per non-poor person targeted														
		National				Intl. 2011 PPP				Percentile-based lines						
		Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
<=17	5.1	0.5:1	3.0:1	18.9:1	29.9:1	18.6:1	All poor	All poor	All poor	1.0:1	1.9:1	7.8:1	18.6:1	29.9:1	All poor	All poor
<=21	9.9	0.4:1	1.9:1	13.9:1	39.4:1	10.2:1	181.0:1	All poor	All poor	0.6:1	1.3:1	4.8:1	11.2:1	22.4:1	326.4:1	961.4:1
<=24	14.9	0.3:1	1.7:1	10.4:1	33.0:1	8.2:1	83.0:1	All poor	All poor	0.5:1	1.1:1	4.3:1	8.9:1	17.3:1	96.3:1	1,438.1:1
<=26	19.5	0.3:1	1.4:1	8.9:1	26.9:1	7.0:1	68.5:1	913.5:1	913.5:1	0.4:1	1.0:1	3.7:1	7.7:1	13.4:1	75.1:1	615.1:1
<=28	24.7	0.3:1	1.3:1	8.2:1	25.5:1	5.9:1	49.0:1	270.8:1	1,154.7:1	0.4:1	0.8:1	3.2:1	6.7:1	12.0:1	63.0:1	210.3:1
<=30	29.2	0.2:1	1.1:1	6.6:1	19.2:1	5.0:1	36.1:1	320.8:1	1,367.4:1	0.3:1	0.8:1	2.8:1	5.6:1	9.5:1	41.9:1	249.2:1
<=32	33.8	0.2:1	1.1:1	6.1:1	19.3:1	4.5:1	37.5:1	342.5:1	1,585.7:1	0.3:1	0.7:1	2.6:1	5.0:1	8.9:1	44.5:1	270.9:1
<=34	39.5	0.2:1	0.9:1	5.2:1	14.4:1	3.9:1	30.0:1	400.2:1	1,852.2:1	0.3:1	0.6:1	2.2:1	4.2:1	7.2:1	36.4:1	316.5:1
<=36	43.8	0.2:1	0.9:1	4.8:1	14.0:1	3.5:1	28.1:1	305.2:1	2,054.5:1	0.3:1	0.6:1	2.1:1	3.8:1	6.8:1	34.8:1	185.3:1
<=38	48.4	0.2:1	0.8:1	4.3:1	12.2:1	3.3:1	23.0:1	156.2:1	2,269.8:1	0.2:1	0.6:1	1.9:1	3.5:1	6.1:1	27.6:1	108.5:1
<=40	53.6	0.2:1	0.7:1	3.6:1	9.7:1	2.8:1	17.8:1	111.7:1	2,512.4:1	0.2:1	0.5:1	1.7:1	3.0:1	5.2:1	21.0:1	79.7:1
<=42	58.0	0.2:1	0.7:1	3.2:1	8.6:1	2.5:1	16.1:1	100.3:1	2,719.6:1	0.2:1	0.5:1	1.6:1	2.7:1	4.5:1	19.3:1	74.4:1
<=45	63.9	0.1:1	0.6:1	2.9:1	7.4:1	2.3:1	13.5:1	78.1:1	1,439.8:1	0.2:1	0.4:1	1.4:1	2.4:1	4.0:1	15.8:1	61.3:1
<=48	68.4	0.1:1	0.6:1	2.5:1	6.4:1	2.0:1	10.9:1	71.6:1	1,541.6:1	0.2:1	0.4:1	1.3:1	2.1:1	3.5:1	13.7:1	53.1:1
<=51	74.0	0.1:1	0.5:1	2.2:1	5.4:1	1.7:1	8.5:1	62.7:1	1,669.5:1	0.2:1	0.4:1	1.1:1	1.8:1	3.0:1	11.1:1	46.2:1
<=54	78.9	0.1:1	0.5:1	1.9:1	4.7:1	1.5:1	7.0:1	47.3:1	1,778.8:1	0.1:1	0.3:1	1.0:1	1.6:1	2.6:1	9.0:1	35.0:1
<=57	83.6	0.1:1	0.5:1	1.7:1	4.1:1	1.4:1	6.0:1	36.7:1	1,886.0:1	0.1:1	0.3:1	0.9:1	1.4:1	2.3:1	7.4:1	26.3:1
<=61	88.2	0.1:1	0.4:1	1.5:1	3.6:1	1.3:1	5.1:1	28.3:1	1,551.9:1	0.1:1	0.3:1	0.8:1	1.3:1	2.0:1	6.3:1	20.2:1
<=66	93.9	0.1:1	0.4:1	1.4:1	3.0:1	1.1:1	4.1:1	20.0:1	647.9:1	0.1:1	0.3:1	0.8:1	1.2:1	1.8:1	5.1:1	15.0:1
<=100	100.0	0.1:1	0.4:1	1.2:1	2.5:1	1.0:1	3.3:1	12.9:1	375.7:1	0.1:1	0.3:1	0.7:1	1.0:1	1.5:1	4.0:1	10.2:1

Scorecard applied to the validation sample. "All poor" means "Only poor targeted".

Figure 21: Share of poor people who are targeted

Targeting cut-off	% all people who are targeted	% poor people who are targeted														
		National				Intl. 2011 PPP				Percentile-based lines						
		Food	100%	150%	200%	$1.90	$3.20	$5.50	$21.70	10th	20th	40th	50th	60th	80th	90th
<=17	5.1	20.5	14.2	8.9	6.9	9.7	6.6	5.5	5.1	25.1	16.6	11.0	9.6	8.2	6.4	5.6
<=21	9.9	32.6	24.4	17.0	13.6	18.2	12.9	10.7	10.0	36.2	28.2	20.1	18.0	15.8	12.4	10.9
<=24	14.9	45.8	34.9	24.9	20.2	26.7	19.1	16.0	14.9	47.7	38.7	29.5	26.4	23.4	18.4	16.3
<=26	19.5	57.0	42.9	32.2	26.3	34.4	25.0	21.0	19.5	57.4	47.6	37.5	34.1	30.2	24.1	21.4
<=28	24.7	63.3	51.4	40.4	33.2	42.5	31.5	26.5	24.7	63.9	55.6	45.8	42.4	37.9	30.4	26.9
<=30	29.2	70.9	57.9	46.6	38.9	49.0	37.0	31.4	29.2	70.7	62.2	52.8	48.9	44.0	35.7	31.9
<=32	33.8	76.1	64.8	53.5	45.1	55.9	42.9	36.4	33.9	76.4	70.4	59.9	55.8	50.7	41.5	37.0
<=34	39.5	81.8	71.0	61.0	51.8	63.2	49.8	42.5	39.6	83.3	76.3	66.5	63.1	57.8	48.2	43.3
<=36	43.8	89.2	76.2	66.8	57.3	68.8	55.1	47.1	43.9	89.1	81.7	72.5	68.5	63.6	53.4	47.9
<=38	48.4	92.2	80.3	72.2	62.7	74.8	60.5	51.9	48.5	91.9	85.6	78.1	74.5	69.3	58.5	52.7
<=40	53.6	96.5	85.0	77.3	68.1	79.8	66.1	57.3	53.7	95.8	91.1	83.0	79.5	74.8	64.1	58.1
<=42	58.0	97.7	89.1	81.3	72.8	83.9	71.2	61.9	58.2	97.4	92.8	87.1	83.6	79.2	69.1	62.9
<=45	63.9	98.7	92.6	87.0	78.8	89.2	77.4	68.0	64.0	98.7	95.3	91.2	88.8	85.0	75.2	69.0
<=48	68.4	98.7	93.9	90.0	82.8	91.6	81.5	72.7	68.5	99.1	96.5	93.0	91.4	88.5	79.8	73.7
<=51	74.0	98.9	96.0	93.3	87.5	94.6	86.3	78.5	74.2	99.3	97.7	95.5	94.4	92.3	85.1	79.6
<=54	78.9	99.5	97.5	95.3	90.9	96.4	89.9	83.3	79.0	99.3	98.4	97.1	96.2	94.6	88.9	84.2
<=57	83.6	99.5	98.4	97.2	94.1	97.8	93.3	87.7	83.8	99.6	99.2	98.3	97.7	96.8	92.3	88.5
<=61	88.2	100.0	98.9	98.4	96.6	98.9	95.9	91.8	88.4	100.0	99.7	98.9	98.8	98.4	95.4	92.3
<=66	93.9	100.0	99.9	99.6	98.8	99.8	98.4	96.4	94.0	100.0	100.0	99.9	99.8	99.6	98.2	96.6
<=100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Scorecard applied to the validation sample.

For example, a pro-poor program could set a score cut-off to achieve a desired poverty rate—say, 80 percent—among targeted people. For 150% of the national line, targeting Tanzanians who score 38 or less would target 48.4 percent of people in Tanzania and give a head-count poverty rate among those targeted of 81.1 percent (Figure 19).

Figure 20 is a different way of looking at this same aspect of targeting accuracy. It shows the number of poor people correctly targeted (included) for each non-poor person mistakenly included (leakage). For 150% of the national line and a score cut-off of 32 or less, 6.1 poor people are successfully targeted for every one non-poor person mistakenly targeted.

Alternatively, a pro-poor program might seek to target a desired share—such as half—of poor Tanzanians. Figure 21 shows that a score cut-off of 32 or less would target 33.8 percent of all Tanzanians, a group in which 53.5 percent are poor by 150% of the national line.

Interview Guide

The excerpts quoted here are from:

NBS (2017) “Instruction Manual: Household Budget Survey, 2017/18”, [the *Manual*].

NBS. (2017) “Questionnaire: Household Budget Survey, 2017/18”, [the *Questionnaire*], link.

G1. Basic interview instructions

The scorecard can be filled out on paper in the field, with responses entered later in a spreadsheet or in your own database. Alternatively, Scorocs’ cloud-based data-collection tool works in a web browser or as an app on Android phones, allowing data entry in the field or in the office. If there is no connection, then data is stored on the phone until it can be uploaded. Try the data-collection tool for the new Tanzania scorecard, or ask about a private account.

The scorecard should be administered by an enumerator trained to follow this “Guide”.

Fill out the scorecard header and the “Back-page Worksheet” first, following the directions on the “Back-page Worksheet”.

In the scorecard header, fill in the number of household members in the space “Number of household members:” based on the list that you the enumerator made as part of the “Back-page Worksheet”.

Do not directly ask the first scorecard question (“In which region does the household live?”). Instead, fill in the response based on your knowledge of the region in which the household lives.

In the same way, do not directly ask the second scorecard question (“How many members does the household have?”). Instead, mark the response based on the number of household members that you listed on the “Back-page Worksheet”.

Ask all of the eight remaining questions directly of the respondent.

Study this “Guide” carefully, and carry it with you while you work. Follow its instructions (including this one).

Remember that the respondent for the interview need not be the household member who is the direct participant with your program.

Likewise, the field agent to be recorded in the scorecard header is not necessarily the same as you the enumerator who does the interview. Rather, the field agent is the employee of the pro-poor program with whom the direct participant has an on-going relationship. If there is no such field agent, then write “NONE” in those spaces in the scorecard header.

In general, do not leave spaces in the header blank. If the requested information is unknown, does not exist, or is not applicable, then write “NONE”, “UNKNOWN”, “DOES NOT EXIST”, or “NOT APPLICABLE” in the blanks. This shows that you the enumerator tried to obtain the data and thus may help avoid the need to return to the household later to try to get the data.

Read each question aloud word-for-word, in the order presented in the scorecard.

When you mark a response to a scorecard question, write the point value in the “Score” column and then circle the spelled-out response option, the pre-printed point value, and the hand-written points, like this:

3. What is the main building material of the floor?	A. Earth/sand, dung, vinyl or asphalt strips, wood planks, palm/bamboo, or other	0	0
	B. Cement, ceramic tiles, or parquet or polished wood	6	

When an issue comes up that is not addressed in this “Guide”, its resolution should be left to the unaided judgment of the enumerator and the respondent, as that apparently was the practice of Tanzania’s NBS in the 2018 HBS. That is, a program should not promulgate any definitions or rules (other than those in this “Guide”) to be used by all its enumerators. Anything not explicitly addressed in this “Guide” is to be left to the unaided judgment of each individual enumerator and the respondent.

Do not read the response options to the respondent. Instead, read the question, and then stop; wait for a response. If the respondent asks for clarification or otherwise hesitates or seems confused, then read the question again or provide additional assistance based on this “Guide” or as you the enumerator deem appropriate.

In general, you should accept the responses given by the respondent. Nevertheless, if the respondent says something—or if you see or sense something—that suggests that the response may not be accurate, that the respondent is uncertain, or that the respondent desires assistance in figuring out how to respond, then you should read the question again and provide whatever help you deem appropriate based on this “Guide”.

While responses to questions in the scorecard are verifiable, in most cases you do not need to verify responses. You should verify only if something suggests to you that a response may be inaccurate and thus that verification might improve data quality. For example, you might choose to verify if the respondent hesitates, seems nervous, or otherwise gives signals that he/she may be lying, confused, or uncertain. Likewise, verification may be called for if a child in the interviewed household or if a neighbor says something that does not square with a respondent’s response. Verification may also be a good idea if you can see something yourself that suggests that a response may be inaccurate, such as a consumer durable that the respondent claims not to possess, or a child eating in the room who has not been counted as a member of the household.

In general, the application of the scorecard should mimic as closely as possible the application of the 2018 HBS by Tanzania’s NBS. For example, interviews should done in-person by a trained enumerator at the residence of the participating household because that is what the NBS did in the 2018 HBS.

G2. Translation

The scorecard itself, the "Back-page Worksheet", and this "Guide" are available in English and Kiswahili. There are not yet official, professional translations to other languages spoken in Tanzania. Users should check scorocs.com to see what translations have been done since this writing. If there is not yet an official, professional translation to a desired language, then please contact Scorocs.

G3. General interview guidance from the *Manual*

G3.1 Meeting the household for the first time

According to p. 7 of the *Manual*, "Your behavior and attitude is vital. As an enumerator, your first responsibility is to establish a good rapport with the interviewed household. At the beginning of an interview, you and the respondent are just beginning to get to know each other. The respondent's first impression of you will influence his/her willingness to cooperate. Be friendly as you introduce yourself.

"Carry a letter and an identification card that indicates that you work with [your organization].

"Wear respectable clothes. Because some households do not have chairs, women are to have *khanga* or *kitenge* for protecting them when sitting down."

G3.2 Your role as enumerator

According to pp. 7–8 of the *Manual*, "You the enumerator play the leading role in the survey. As an enumerator, your duty is to collect the required data. The quality of your work is determined by the quality of the data that you collect. The success or failure of the survey is in your hands.

"Carefully follow the instructions in this *Manual* [including this one].

"Create a friendly atmosphere of trust with the respondent so that you can get accurate data. To this end, it helps to familiarize yourself with the culture and habits of the community.

"The first impression that you make by your appearance, first words, and actions are vital for winning the respondent's cooperation.

"Demonstrate to respondents that you know what you are doing. Let them know that their participation will benefit [the participants in your program].

"Do not promise any immediate benefit to the interviewed household in return for its participation, as this may reduce the accuracy of its responses.

“Highlight that responses will be kept strictly confidential.

“Do not allow third parties to be present for the interview. You the enumerator should be alone with the members of the interviewed household.

“Be efficient in your work, but do not rush the respondent.

“Unlike a normal conversation, you the enumerator ask all the questions in the interview, and the respondent gives all the responses. Keep your opinions to yourself, and do not react in any way to what the respondent says. Always remain neutral, and do not show any approval nor disapproval.

“If the respondent is reluctant or unwilling to respond to a question, then try to coax him/her to respond. Explain again that all information is confidential. But do not push too hard; a respondent cannot be forced to respond.

“Strictly follow the sequence of questions as written.” Read them word-for-word.

“Take charge of the interview. Maintain the interest of the respondent at all times. If the respondent digresses or rambles on, then do not interrupt him/her abruptly or rudely. Listen to what he/she has to say, and then gently try to steer him/her back to the question.”

G3.3 Confidentiality

According to p. 4 of the *Manual*, the information that you collect “is strictly confidential and is not to be disclosed to any person except those who are involved in the survey. . . . [Assure] the respondent that his/her information will be treated strictly as confidential and will be used for statistical purposes only.”

G3.4 Who should be the respondent?

Remember that the respondent does not need to be the household member who is the direct participant with your program (although the respondent may be the direct participant).

According to page 6 of the *Manual*, “If the head of the household is not available, then any other member of the household can be interviewed.” This implies that the preferred respondent is the head of the household.

G3.5 Who is the head of the household?

Note that the head of the household may or may not be the household member who is the direct participant with your program (although the head may be the direct participant).

According to p. 13 of the *Manual*, the *head of the household* is "the person recognized as the head by the other members of the household. Often he/she is responsible for the financial support and welfare of the household members."

G4. Guidelines for each question in the scorecard

G4.1 In which region does the household live?

A. Ruvuma, or Lindi
B. Dodoma, Mwanza, Rukwa, or Shinyanga
C. Tanga, Morogoro, Arusha, or Iringa
D. Pwani, Geita, Simiyu, Manyara, Njombe, or Katavi
E. Kigoma, Mtwara, Singida, Tabora, or Songwe
F. Dar es Salaam, Kagera, Mbeya, Kilimanjaro, or Mara

Unless you have to, do not directly ask this question of the respondent. Instead, fill in the response based on your knowledge of the region in which the household lives.

G4.2 How many members does the household have?

A. Eight or more
B. Seven
C. Six
D. Five
E. Four
F. Three
G. Two
H. One

Do not directly ask this question of the respondent. Instead, mark the response based on the number of household members that you listed on the "Back-page Worksheet".

According to p. 13 of the *Manual*, a *household* is made up of members who share consumption and who pool their resources together.

"A household may consist have one member, or more than one member.

"A *one-person household* consists of a person who lives alone, occupying all or part of a residence, and who has independent consumption.

"A *multi-person household* is a group of two or more people who together who occupy all or part of a residence and who share their consumption. A common type of household is made up of a husband, a wife, and their children. Other relatives, boarders, visitors, and other people are counted as *members of the interviewed household* if they pool resources, share consumption, and have lived with the interviewed household for at least two weeks.

"Domestic servants count as members of the interviewed household if and only if they eat their meals with the interviewed household and recognize the head of the interviewed household as their head (regardless of whether the domestic servants sleep in the residence of the interviewed household).

"Children of members of the interviewed household who are at boarding school count as members of the interviewed household.

"Lodgers/boarders/residents who pay the interviewed household for shelter and meals count as members of the interviewed household. Others who reside with the interviewed household but do not share resources and meals with the interviewed household do not count as members of the interviewed household.

“A man with wives in more than one household is to be counted as a member of the interviewed household if he spends more than half of his time with the interviewed household.”

Any given person is a member of one (and only one) household. Any given household has one (and only one) head. The head of a given household must be a member of that household. No one can be the head of more than one household.

According to Section 1 (“Respondent’s characteristics”) of the *Questionnaire*, a *household* is “all the people who normally live [in the residence with the direct participant with your program] and who eat their meals together.

When completing the “Back-page Worksheet”, record the names of household members in the following order:

- “The household head
- The spouse/conjugal partner of the head (if there is one)
- Other members of the immediate (nuclear) family of the head who normally live [in the residence with your program’s direct participant] and who eat their meals together
- Other relatives or non-relatives (such as live-in domestic servants) who normally live [in the residence with your program’s direct participant] and who eat their meals together
- Anyone else who is not present now in the residence but who normally lives [in the residence with your program’s direct participant] and who eat their meals together”

G4.3 What is the main building material of the floor?

A. Earth/sand, dung, vinyl or asphalt strips, wood planks, palm/bamboo, or other
B. Cement, ceramic tiles, or parquet or polished wood

According to p. 43 of the *Manual*, you the enumerator should ask this question of the respondent; do not mark a response based solely on what you yourself can see of the floor.

This question pertains to the floor of the main building of the residence of the interviewed household.

Record the type of material that accounts for the largest share of the floor's area.

According to p. 14 of the *Manual*, a *residence* is "the living space occupied by one household, regardless of its physical arrangement. For example, a residence may be a single room, or it may be one (or more than one) housing units."

G4.4 Does the household own any charcoal stoves?

A. No
B. Yes

According to p. 52 of the *Manual*, “Do not count charcoal stoves that are not in good working condition unless you the enumerator verify that the household expects them to be repaired in the next three months.

“Charcoal stoves that the household received as aid should be counted.”

G4.5 Does the household own any tables?

A. No
B. Yes

According to p. 52 of the *Manual*, “Do not count tables that are not in good working condition unless you the enumerator verify that the household expects them to be repaired in the next three months.

“Tables that the household received as aid should be counted.”

G4.6 Does any member of the household own a television?

A. No
B. Yes

According to p. 52 of the *Manual*, “Do not count televisions that are not in good working condition unless you the enumerator verify that the household expects them to be repaired in the next three months.

“Televisions that the household received as aid should be counted.”

G4.7 Does any member of the household own a mobile phone?

A. No
B. Yes

According to p. 52 of the *Manual*, “Do not count mobile phones that are not in good working condition unless you the enumerator verify that the household expects them to be repaired in the next three months.

“Mobile phones that the household received as aid should be counted.”

G4.8 Does any member of the household own any cattle, goats, sheep, or hogs?

A. No
B. Yes

According to p. 70 of the *Manual*, "Count only cattle, goats, sheep, or hogs owned by members of the interviewed household. Do not count cattle, goats, sheep, or hogs that are in the care of the interviewed household but that are owned by some other household."

G4.9 In the past week, did the household consume any chicken, goat, fish, beef, or other meat?

A. No
B. Yes

The *Manual* does not have any additional information about this question.

G4.10 How many meals does the household usually have per day?

A. Two or less
B. Three or more

According to p. 74 of the *Manual*, “This question concerns the number of main meals that the interviewed household usually eats per day. This includes breakfast, lunch, and dinner. These are meals prepared and taken at home.”

Technical Annexes: Overview

The technical annexes cover aspects of the scorecard for advanced users or specialists. While program managers can skip the annexes and still benefit from using the scorecard, understanding the details will increase the usefulness of scorecard estimates and improve implementation, interpretation, and analysis.

The annexes cover:

Annex 1 Data used for construction and validation

Tanzania's National Bureau of Statistics (NBS) fielded the 2018 Household Budget Survey (HBS) with 9,462 households from December 1, 2017 to November 30, 2018. The 2018 HBS is Tanzania's most-recent national household consumption survey.

Questions and points for the scorecard are selected (*constructed*) based on data from a random three-fifths of the 9,462 households in the 2018 HBS. These same three-fifths of households are also used to associate (*calibrate*) scores with poverty likelihoods for all poverty lines.

Data from the other two-fifths of households from the 2018 HBS is used to test (*validate*) the scorecard's accuracy for one-period, snapshot estimates of poverty rates *out-of-sample*, that is, with data that is not used in construction nor calibration. Data from those same two-fifths of households are also used for out-of-sample validation of targeting accuracy.

Annex 2 Definition of *poverty*

A household's *poverty status* as poor or non-poor depends on whether its consumption expenditure (TZS per person per day or per adult equivalent per day) is below a given poverty line. Thus, a definition of *poverty* is a poverty line together with a measure of consumption expenditure.

NBS (2019, p. 7) documents Tanzania's definition of *consumption expenditure*.

Because pro-poor programs in Tanzania may want to use different or various poverty lines, the scorecard supports 15 lines:

- Food (extreme) line
- 100% of the national (basic-needs) line
- 150% of the national line
- 200% of the national line
- $1.90/day 2011 PPP
- $3.20/day 2011 PPP
- $5.50/day 2011 PPP
- $21.70/day 2011 PPP
- First-decile (10^{th}-percentile) line
- First-quintile (20^{th}-percentile) line
- Second-quintile (40^{th}-percentile) line
- Median (50^{th}-percentile) line
- Third-quintile (60^{th}-percentile) line
- Fourth-quintile (80^{th}-percentile) line
- Tenth-decile (90^{th}-percentile) line

A2.1 National poverty lines

Tanzania's official national poverty lines are defined in terms of minimum standards for food and non-food consumption.

The food (extreme) poverty line is the minimum food standard. This is the cost of 2,200 Calories from a basket of the 251 most-common food items, with shares based on people in the 2018 NBS in the 10th to 50th percentiles of per-adult-equivalent food consumption.[36] The food line is adjusted for differences in prices across the 104 combinations of 26 regions and four quarters during the 2018 fieldwork. In average prices for Tanzania overall during the 2018 HBS, the food line is TZS1,108 per adult equivalent per day, giving a head-count poverty rate of 8.0 percent (Figure 10).

"100% of the national (basic-needs) line" is the sum of the food and non-food standards. If the non-food standard uses the same cost-of-basic-needs method[37] as in the 2011/12 HBS,[38] then it is the average non-food consumption of people whose total per-adult-equivalent consumption is between the food line and 120 percent of the food line, adjusted for price differences across 104 region-quarters. For Tanzania overall, this is TZS1,619 per adult equivalent per day, with a head-count poverty rate of 26.4 percent.[39]

150% of the national line and 200% of the national line) are multiples of 100% of the national line.

[36] NBS, 2019.
[37] Ravallion, 1998.
[38] World Bank, 2015.
[39] This rate matches NBS (2019, p. 10), suggesting that this paper uses the same data and calculations as NBS did.

The 2018 poverty-line definitions that are supported by the new scorecard here are not comparable with those used for Tanzania's old scorecards based on data from 2007 and 2011/12 (Schreiner, (2016a and 2013a).[40] As a result, estimates from an old scorecard are not comparable with estimates from the new scorecard. Thus, estimates of changes in poverty over time must use the same scorecard for both baseline and follow-up. Do not use an old scorecard at baseline and the new scorecard at follow-up.

A2.2 International 2011 PPP poverty lines

The World Bank tracks world-wide poverty with four poverty lines:[41]

- $1.90/day Low-income countries (the international "extreme poverty" line)
- $3.20/day Lower-middle-income countries
- $5.50/day Upper-middle-income countries
- $21.70/day High-income countries

Although the World Bank classifies Tanzania as an lower-middle income country, the most relevant international line is $1.90/day, as about half of Tanzanians are below that line.

PPP lines control for differences in purchasing power across countries due to the fact that non-tradable goods and services are usually less costly in poorer countries while tradables are more costly. PPP adjustments increase the comparability of poverty estimates across countries.

---

[40] World Bank (2015) documents how the definitions differ from 2007 to 2011/12, and NBS (2014) eschews—and repeatedly warns against—comparing poverty estimates across those rounds. The 2018 definition differs from 2011/12 in that it:

- Takes its sampling frame from the 2012 Household and Population Census
- Stratifies by region (rather than Dar es Salaam, rural, and other urban)
- Has a 14-day consumption diary (rather than 18 days)
- Adjusts prices across 104 area-periods (rather than 12)

Nevertheless, NBS (2019, pp. 11 and 28) compares poverty rates across 2007, 2011/12, and 2018 (as well as 2000/1 and 1991/2).

[41] Jolliffe and Prydz, 2016; Ferreira *et al.*, 2016.

International 2011 PPP lines for Tanzania are derived from:

- 2011 PPP (revised) exchange rate for Tanzania for "individual consumption expenditure by households":[42] TZS588.785 per $1.00
- Average all-Tanzania Consumer Price Index[43] (CPI):
  - Calendar-year 2011: 112.69
  - December 1, 2018, to November 30, 2019: 180.68
- Average all-Tanzania spatial and temporal price deflator during the 2018 HBS fieldwork: 0.9981739
- Region-quarter price deflators for 100% of the national line from HBS data

Given this, the $1.90/day 2011 PPP line for a given household is:

$$\$5.50 \cdot 2011 \text{ PPI factor} \cdot \frac{\text{Deflator}_{\text{Household}}}{\text{Ave. deflator}_{\text{Tanzania}}} \cdot \frac{\text{CPI}_{\text{2018HBS}}}{\text{CPI}_{\text{2011}}}.$$

For example, the deflator for a household in Songwe in the second quarter of the 2018 HBS is 0.9797897. The $1.90 line in that region-quarter in prices for Tanzania overall on average during the 2018 HBS fieldwork is then:

$$\$1.90 \cdot 588.785 \cdot \frac{0.9797897}{0.9981739} \cdot \frac{180.68}{112.69} = \text{TZS1,761}.$$

For Tanzania overall, the average $1.90/day line is TZS1,794 per person per day, with a head-count poverty rate of 48.9 percent.

The 2011 PPP lines for $3.20/day, $5.50/day, and $21.70/day are multiples of the $1.90/day line.

For the 2018 HBS, the World Bank's PovcalNet reports the same $1.90/day line of TZS1,794.[44] PovcalNet's head-count poverty rate, however, is 0.5 percentage points higher because PovcalNet incorrectly takes the average deflator as 1.00. The rate reported here is to be preferred.

[42] World Bank, 2020, Table E.3, column 13, p. 134.
[43] Base = 100 in calendar-year 2010, link.
[44] PovcalNet, 2020.

A2.3 Percentile-based poverty lines

The scorecard also supports percentile-based poverty lines.[45] This facilitates a number of types of analyses. For example, the second-quintile (40th-percentile) line might be used to help track Tanzania's progress toward the World Bank's (2013) goal of "shared prosperity/inclusive economic growth", defined as income growth among the bottom 40 percent of the world's people.

The four quintile lines (or all seven percentile lines), analyzed together, can also be used to look at the relationship of consumption expenditure with health outcomes (or anything else related with the distribution of consumption expenditure). The scorecard thus offers an alternative for health-equity analyses that typically have used an asset index (such as that supplied with the data from the Demographic and Health Surveys) to compare an estimate of socio-economic status with health outcomes.[46]

Of course, relative-wealth analyses are also possible with scores from the scorecard. But support for relative consumption-expenditure lines allows for a more straightforward use of a single tool to analyze any or all of:

- Relative wealth (via scores)
- Absolute consumption expenditure (via poverty likelihoods and absolute poverty lines)
- Relative consumption expenditure (via poverty likelihoods and percentile-based poverty lines)

Unlike the scorecard, asset indexes only estimate relative wealth. Furthermore, the scorecard—unlike asset indexes—uses a straightforward, well-understood standard for socio-economic status whose definition is external to the tool itself (consumption expenditure relative to a poverty line defined in monetary units).

In contrast, an asset index defines *poverty* in terms of its own questions and points, without calibration or reference to an external standard. This means that two asset indexes with different questions or different points—even if derived from the same data for a given country—imply two distinct definitions of *poverty*. In the same set-up, two scorecards would provide comparable estimates under a single definition of *poverty*.

---

[45] Percentiles are defined in terms of all people in Tanzania. For example, the all-Tanzania head-count poverty rate for the first-quintile (20th-percentile) poverty line is 19.9 percent (Figure 10).

[46] Rutstein and Johnson, 2004.

Annex 3 Scorecard construction

For Tanzania, about 70 candidate questions are prepared in these areas:

- Household composition (such as the number of household members)
- Education (such as the highest level completed by the head of the household))
- Housing (such as the main material of the floor)
- Ownership of consumer durables (such as charcoal stoves or televisions)
- Recent food consumption (such as whether the household ate any chicken, goat, fish, beef, or other meat in the past week)
- Location of residence (such as the region)

To facilitate the estimation of change over time, preference is given to questions that are more sensitive to changes in poverty. For example, the consumption of meat is probably more responsive to changes in poverty than is the age of the head of the household).

The scorecard itself is built using 150% of the national poverty line and Logit regression on the construction sub-sample. Questions are selected based on both judgment and statistics.

The first step is to use Logit to build a draft scorecard for each candidate question. The power of each one-question draft scorecard to rank households by poverty status is assessed via the concentration index.[47]

[47] Ravallion, 2009.

One of the one-question draft scorecards is then selected based on:[48]

- Improvement in accuracy
- Acceptability to users in terms of:
  - Simplicity
  - Cost of collection
  - Concordance with:
    - Experience
    - Theory
    - Common sense
- Sensitivity to changes in consumption expenditure
- Variety among types of questions
- Applicability across regions
- Tendency to have a slow-changing relationship with poverty
- Relevance for distinguishing among people at the poorer end of the distribution of consumption expenditure
- Verifiability

A series of two-question draft scorecards are then built, each adding a second question to the one-question scorecard selected from the first stage. The best two-question draft scorecard is then selected, again using judgment to balance statistical accuracy with non-statistical criteria. These steps are repeated until the scorecard has 10 questions that work well together.

The last step is to transform the Logit coefficients into non-negative integers such that scores range from 0 to 100, with lower scores corresponding with greater poverty.

This algorithm is similar to common R^2-based stepwise least-squares regression. It differs from naïve stepwise in that the selection of questions considers both statistical[49] and non-statistical criteria. The use of non-statistical criteria can improve robustness through time and across non-nationally representative groups. It also helps to ensure that questions are straightforward, common-sense, inexpensive-to-collect, and acceptable to users.

---

[48] Schreiner *et al.*, 2014; Zeller, 2004.

[49] The statistical criterion is not the *p* values of coefficients but rather a question's contribution to the ranking of households by poverty status in the context of a scorecard with nine other questions.

The single scorecard here applies to all of Tanzania. Customizing poverty-assessment tools by urban/rural does not improve targeting accuracy much.[50] Segment-specific tools may improve the accuracy of estimates of poverty rates,[51] but:

- They run a greater risk of overfitting[52]
- Most of their benefit can be had in a single scorecard that includes a question that identifies the specific segment of interest (such as, in the case of Tanzania, the region of residence)[53]

[50] Brown, Ravallion, and van de Walle, 2018; World Bank, 2012; Sharif, 2009; Schreiner, 2006; Schreiner, 2005; Narayan and Yoshida, 2005; and Grosh and Baker, 1995.
[51] Diamond *et al.*, 2016; Tarozzi and Deaton, 2009.
[52] Haslett, 2012.
[53] Schreiner, 2016c.

Annex 4 Estimates of poverty likelihoods

This annex tells how scores are converted into estimated poverty likelihoods.

Scores are on an ordinal scale from 0 to 100. Higher scores signal less poverty, but not how much less. The ordered symbols used to represent scores are numbers, but those symbols are not the normal cardinal numbers that you can do math on. For example, a score of 20 plus a score of 10 is not 30 of anything, just as the letter "A" plus the letter "B" is not the letter "C" (nor anything else).

To get cardinal units, a look-up table is used to convert scores to *poverty likelihoods*, that is, probabilities of being below a poverty line. For the example of 150% of the national line, scores of 31–32 correspond with a poverty likelihood of 73.3 percent, and scores of 33–34 correspond with a poverty likelihood of 72.8 percent (Figure 1).

The poverty likelihood associated with a score varies by poverty line. For example, scores of 31–32 are associated with a likelihood of 73.3 percent for 150% of the national line but with a likelihood of 69.3 percent for the $1.90/day 2011 PPP line.

A4.1 Calibrating scores with poverty likelihoods

A given score is associated ("calibrated") with an estimated poverty likelihood that is defined as the share of people in the construction sub-sample who have the score and who live in households with per-capita or per-adult-equivalent consumption expenditure below a given poverty line.

For the example of 150% of the national line and a score of 31–32 (Figure 22 below), there are 3,562 (normalized) households in the construction sample. Of these, 2,612 (normalized) are below the poverty line. The estimated poverty likelihood associated with a score of 31–32 is then 73.3 percent, because 2,612 ÷ 3,562 ≈ 0.733 = 73.3 percent.

The same method is used to calibrate all scores with poverty likelihoods for all 15 poverty lines.[54]

[54] To ensure that likelihoods never increase as scores increase, likelihoods across adjacent scores may be non-parametrically smoothed before grouping scores into ranges. This preserves unbiasedness while preventing higher scores from being associated with higher likelihoods.

Figure 22: Estimation of poverty likelihoods (150% of national line)

Score	Households in range and < poverty line		All households in range		Poverty likelihood (%)
0–17	2,723	÷	2,873	=	94.8
18–21	2,686	÷	2,990	=	89.8
22–24	2,961	÷	3,424	=	86.5
25–26	2,474	÷	3,022	=	81.9
27–28	2,771	÷	3,418	=	81.1
29–30	2,952	÷	3,778	=	78.1
31–32	2,612	÷	3,562	=	73.3
33–34	3,373	÷	4,635	=	72.8
35–36	2,829	÷	4,157	=	68.0
37–38	2,350	÷	3,732	=	63.0
39–40	2,776	÷	4,777	=	58.1
41–42	2,315	÷	4,129	=	56.1
43–45	2,907	÷	5,490	=	53.0
46–48	2,605	÷	6,049	=	43.1
49–51	1,834	÷	5,617	=	32.7
52–54	1,450	÷	5,677	=	25.5
55–57	1,128	÷	5,099	=	22.1
58–61	844	÷	6,037	=	14.0
62–66	685	÷	7,718	=	8.9
67–100	347	÷	13,817	=	2.5

Number of all households normalized to sum to 100,000.

A4.2 Objectivity of estimates of poverty likelihoods

Even though scorecard questions are selected partly based on judgment related to non-statistical criteria, the calibration process produces estimates of poverty likelihoods that are objective, that is, derived from monetary poverty lines and from survey data on consumption expenditure.[55] The fact that some choices in scorecard construction are informed by judgment in no way impugns the objectivity of the estimated likelihoods; their objectivity depends on using data in score calibration, not on using data (and nothing else) in scorecard construction.

A4.3 Why not use the Logit formula?

The scorecard is based on a Logit regression (Annex 3). This means that poverty likelihoods could be estimated not with a calibrated look-up table (Figure 1) but rather with the Logit formula of $2.718281828^{\beta X}$ x $(1 + 2.718281828^{\beta X})^{-1}$, where β is a vector of the Logit coefficients and X is a vector of a household's responses.

The scorecard uses the calibration approach is because the Logit formula looks scary. Program managers can understand poverty likelihoods defined as the share of people with a given score in the construction sample from Tanzania's 2018 HBS who are below a poverty line. A calibrated look-up table also allows program analysts to convert scores to likelihoods without any math at all. This calibration approach can also improve accuracy, especially with large samples.

[55] The calibrated likelihoods would be objective even if scorecard construction did not use any data at all. In fact, objective scorecards of proven accuracy are often constructed using only expert judgment (Caire, 2004; Schreiner *et al.*, 2014).

Annex 5 Error and margins of error

This annex reports the scorecard's estimation error for head-count poverty rates in a single time period. It also discusses margins of error for all estimates.

A5.1 Estimation errors

A5.1.1 What is estimation error?

Estimation error is the distance and direction by which a scorecard's estimate tends to miss the true value in the population.

For example, the estimation error of Tanzania's scorecard for snapshot estimates of head-count poverty rates in a single time period by 150% of the national poverty line is –1.8 percentage points (Figure 2).

An unadjusted estimate can usually be improved—that is, moved closer to the true value—by subtracting off the known estimation error. For example, if the unadjusted estimate is 90.2 percent and the error is –1.8 percentage points, then an improved estimate is 90.2 – (–1.8) = 92.0 percent.

A5.1.2 What estimation errors are reported here?

Estimation errors are reported for snapshot estimates of head-count poverty rates in a single time period for all 15 poverty lines.

Errors are derived *out-of-sample*. This means that the scorecard (made from the construction sample from the 2018 HBS, Annex 1) is tested with repeated sub-samples from the validation sample that were not used to construct the scorecard. The estimation error is the average of the differences between scorecard estimates and observed poverty rates across these repeated sub-samples.

There is no data today on consumption-based poverty in the future, so it is impossible to report estimation error for annual net changes in head-count poverty rates across two time periods. The scorecard cannot be not tested *out-of-time* because it is both constructed and validated with data from a single time period (2018).

In practice, the scorecard—like all poverty-assessment tools—is always applied both out-of-sample and out-of-time. Being out-of-sample violates the assumption that the scorecard is applied to a sample from the same population whose data was used to construct the scorecard. Being out-of-time violates the assumption that the relationships between poverty and scorecard questions are the same as in the population whose data was used to construct the scorecard.

The unknown degree of these inevitable violations of the scorecard's assumptions means that actual estimation errors will differ from those reported here in unknowable ways.[56] Still, the errors (and margins of error) reported here are the best available, and it makes sense to account for them.

A5.1.3 How to estimate estimation errors

Given the scorecard's standard assumptions, an unbiased estimator of *estimation error* is the average of differences between scorecard estimates and observed values in repeated sub-samples from the validation sample.[57]

It is possible to compare estimated and observed poverty rates because the validation sample from the 2018 HBS records actual (not estimated) consumption-based poverty status. The observed poverty likelihood in the 2018 HBS is 100 percent for poor households and 0 percent for non-poor households. For a given poverty line, the observed (not estimated) head-count poverty rate is the household-size-weighted average of observed poverty likelihoods.

The scorecard can also be applied to the same validation sub-sample (ignoring that actual poverty status is observed) to estimate the poverty rate as the household-size-weighted average of estimated poverty likelihoods (Section 3).

The scorecard's estimation error in a given validation sub-sample is then the difference between the scorecard estimate versus the observed value.

---

[56] Estimation errors due to being out-of-time can be measured with post-2018 data (say, from a future HBS). Of course, future HBS data is not yet available, and even after it is available, there will still be some unknown out-of-time error (and out-of-sample error will still be completely unknown).

[57] This is the *bootstrap approach*. The average of estimates from repeated samples from the validation sample is an unbiased estimator of the true value in the population of Tanzania overall. The population's true value is taken as the value in the 2018 HBS (even though the HBS is itself only a sample).

Different sub-samples from the validation sample result in different errors. The estimate of the scorecard's general *estimation error* is the average of these errors across many sub-samples.[58] In turn, the scorecard estimate's margin of error reflects the extent of the spread of the distribution of all the sub-samples' errors around their average.[59]

A5.1.4 Errors for snapshot estimates of poverty rates in one time period

The first line in Figure 2 ("Estimation error") presents errors for snapshot estimates of poverty rates in one time period for Tanzania's 15 poverty lines.

The average of the absolute value of each error across all poverty lines is about 1.9 percentage points. The largest absolute error is 5.2 percentage points. The error for 150% of the national line is –1.8 percentage points.

A5.2 Margins of error

A5.2.1 What are margins of error?

Like any statistic, a scorecard estimate depends on a particular sample from a population. Because samples are drawn at random, each sample is different, and different samples give different scorecard estimates. Scorecard estimates are unbiased—under the standard assumptions—because the average of estimates across many repeated samples is the same as the single true value in the population.

Unusual luck in any single sample, however, may push an estimate for that sample far from the true value in the population. Larger samples provide more chances for luck to even out, so large errors are less likely in larger samples.[60]

---

[58] Households in a sub-sample are drawn *with replacement*; each draw comes from the full pool, including households that have already been drawn. Thus, a given household may appear in a given sub-sample once, more than once, or not at all.

[59] Schreiner (2020) discusses the derivation of errors.

[60] When flipping a fair (unbiased) coin, the true probability of "heads" is 50 percent. *Unbiasedness* means that the average of the share of "heads" in many samples will be close to 50 percent. In a single sample of 10 tosses, however, the chances of at least six "heads" (60 percent of tosses, with an error of at least 10 percentage points) is about 37 percent. In a single sample of 100 tosses, the chances of such a large error is smaller (about 3 percent). Larger samples reduce the risk that estimates will be far from true values.

For a given estimate, sample size, and confidence level, the *margin of error* is the range of true population values that is consistent with the estimate.

A margin of error has two parts:

- The margin of error itself (such as ±2.0 percentage points). This range is centered on the estimate
- A confidence level (such as 90 percent) that the true value falls within the margin of error

All else constant, narrower margins of error or higher confidence levels mean that it is more likely that the sample-based estimate is closer to the true population value.

To illustrate, suppose that the adjusted estimate of the head-count poverty rate for 150% of the national line is 92.0 percent and that the sample size is n = 1,024. Given 90-percent confidence,[61] the margin of error is ±2.9 percentage points (Figure 2). Absent other sources of error and given the scorecard's standard assumptions, this means that there is a 90-percent chance that the true population value is in the range of 92.0 – 2.9 = 89.1 percent to 92.0 + 2.9 = 94.9 percent, with the most-likely true value being the center of the range (the 92.0-percent estimate).

Said another way, "With 90-percent confidence, the estimate has a margin of error from 89.1 to 94.9 percent." This means that the true population value has a:

- 5-percent chance of being less than 89.1 percent
- 90-percent chance of being between 89.1 and 94.9 percent
- 5-percent chance of being greater than 94.9 percent

A5.2.2 Why do margins of error matter?

Managers should put more weight on estimates with narrower margins of error.

As a hypothetical example, a pro-poor program in Tanzania probably is indeed pro-poor if the scorecard estimate of the poverty rate for in-coming participants by 150% of the national poverty line with 80-percent confidence is 65.0 percent with a margin of error of ±5.0 percentage points, that is, from 60.0 to 70.0 percent. The estimate and its margin of error suggest that the true poverty rate of in-coming participants is unlikely to be less than or about the same as the all-Tanzania rate for this line of 54.0 percent from Figure 10.

[61] Most real-world decisions are made with much less than 90-percent confidence.

If, however, the margin of error were ±15.0 percentage points (that is, from 50.0 to 80.0 percent), then there would be a non-negligible chance that the poverty rate of in-coming participants is less than or about the same as that of the average Tanzanian and thus that the program may not actually be pro-poor.

So far, almost all analyses of scorecard estimates have ignored margins of error. This deficient practice increases the risk of bad decisions. Do not make this mistake.

A5.2.3 Margins of error for snapshot estimates of poverty rates in one time period for the Tanzania scorecard

For sample sizes of n = 1,024 and 90-percent confidence and across all supported poverty lines, the margins of error for snapshot estimates of head-count poverty rates in a single time period are ±3.1 percentage points or smaller (Figure 2). Given the scorecard's standard assumptions, this means that in 90 of 100 samples of this size, the true population value is within ±3.1 percentage points or less of the error-adjusted estimate.

A5.2.4 How to calculate margins of error

The spreadsheet-based PovIt!TM reporting app calculates margins of error for all scorecard estimates discussed here. Analysts may also use the formulas below.[62]

[62] Schreiner (2020) discusses the derivation of the formulas.

A5.2.5 Formula for margins of error for snapshot estimates of head-count poverty rates in a single time period

All formulas for margins of error involve the following elements:

$\pm c$ is the margin of error as a proportion (*e.g.*, ±0.020 for ±2.0 percentage points),

z is from the Normal distribution and is $\begin{cases} 1.04 \text{ for confidence levels of 70 percent} \\ 1.28 \text{ for confidence levels of 80 percent} \\ 1.64 \text{ for confidence levels of 90 percent} \end{cases}$,

σ is the standard error of the estimated poverty rate, that is, $\sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \cdot \varphi$,

$\hat{p}$ is the estimated poverty rate as a proportion,

φ is the finite population correction factor $\sqrt{\frac{N-n}{N-1}}$,

N is the population size in terms of households (not members of households),

n is the sample size (in terms of interviewed households,
not members of interviewed households), and

α is an adjustment factor specific to the scorecard, estimator, and poverty line.

Given a confidence level that corresponds with z, a sample-based estimate $\hat{p}$, a population N, a sample n, and an adjustment factor α for a specific poverty line from Figure 2, the formula[63] for the margin of error $\pm c$ is $\pm z \cdot \alpha \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$.

To illustrate, Tanzania's 2018 HBS gives a direct-measure head-count poverty rate for 150% of the national line of $\hat{p}$ = 54.0 percent (Figure 10). The adjustment factor α is 1.00 by definition because $\hat{p}$ is a direct-measure estimate, not an indirect-scorecard estimate. Tanzania in 2018 had a population of households (not people) of N = 11,349,526, and the HBS sample size was n = 9,462. Given a desired confidence level of 90 percent, z is 1.64. The margin of error $\pm c$ is then about ±0.8 percentage points:

$$\pm z \cdot \alpha \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \pm 1.64 \cdot 1.00 \cdot \sqrt{\frac{0.540 \cdot (1-0.540)}{9{,}462}} \cdot \sqrt{\frac{11{,}349{,}526-9{,}462}{11{,}349{,}526-1}}.$$

This implies a 90-percent chance that Tanzania's true head-count poverty rate for 150% of the national line in 2018 is in the range of 54.0 – 0.8 = 53.2 percent to 54.0 + 0.8 = 54.8 percent.

A5.2.6 Margins of error for snapshot estimates of numbers of poor people in a single time period

The lower (upper) limit of the margin of error for a snapshot estimate of numbers of poor people is the number of people in participating households, multiplied by the lower (upper) limit of the margin of error of the poverty-rate estimate.

[63] This formula ignores how sampling variability affects the derivation of the scorecard. It also ignores that household size varies and that larger households are more likely to have higher poverty likelihoods. This understates the margin of error.

To illustrate, the baseline example in Section 3 has an estimated snapshot poverty rate of 92.0 percent. With 70-percent confidence, the margin of error is about ±22.7 percentage points,[64] or from 92.0 – 22.7 = 69.3 percent to 92.0 + 22.7 = 114.7 percent ≈ 100 percent (because a poverty rate cannot exceed 100 percent). The margin of error is huge because the sample size of $n = 2$ interviewed households is very small.[65]

The estimated number of people in participating households in the example in Section 3 is 7,000,[66] so the lower limit of the 70-percent margin of error for the estimated number of poor people is 7,000 · 0.693 = 4,851. The upper limit is 7,000 · 1.00 = 7,000.

A5.2.7 Margins of error for estimates of the annual net change in head-count poverty rates across two periods for one sample, scored twice

In this case, the formula for the margin of error $\pm c$ is:

$$\pm \frac{z \cdot \alpha}{y} \cdot \sqrt{\frac{\hat{p}_{up} \cdot (1 - \hat{p}_{up}) + \hat{p}_{down} \cdot (1 - \hat{p}_{down}) + 2 \cdot \hat{p}_{up} \cdot \hat{p}_{down}}{n}} \cdot \sqrt{\frac{N - n}{N - 1}},$$

where:

- z, α, N, and n are defined as above
- $\hat{p}_{up}$ is the share of members of sampled households that rise above the poverty line from below
- $\hat{p}_{down}$ is the share of members of sampled households that fall below the poverty line from above
- y is the household-size-weighted average of years between interviews

---

[64] The example in Section 3 has $N = 1{,}000$, $n = 2$, and $\alpha = 1.14$. For 70-percent confidence, $z = 1.04$. The margin of error $\pm c$ for the head-count poverty-rate estimate is then $\pm 0.227 \approx \pm 1.04 \cdot 1.14 \cdot \sqrt{\frac{0.920 \cdot (1 - 0.920)}{2}} \cdot \sqrt{\frac{1{,}000 - 2}{1{,}000 - 1}}$.

[65] Yet the formulas for margin of error still apply, and the estimator is still unbiased.

[66] The formula for margin of error for the estimated number of poor people ignores that the estimated number of people in participating households has its own margin of error. This understates the margin of error.

Illustrating with the earlier example of one sample scored twice (Section 3.3.1), $\hat{p}_{up}$ is the number of household members estimated to rise above a poverty line from below. This is the absolute value of the sum of the estimated *negative* changes in the number of members in poor households (from column M in Figure 11, here |–0.706 + (–0.036)| = +0.742), divided by the sum across all sampled households of each household's average household size across baseline and follow-up of 7.0 + 6.0 = 13.0 (from columns E and F). Thus, $\hat{p}_{up}$ = 0.742 ÷ 13 ≈ 0.057.

In turn, $\hat{p}_{down}$ is the share of household members estimated to fall below a poverty line from above. This is the sum of the estimated *positive* net changes in the number of members in poor households (from column M in Figure 11), which is zero in this example because there are no positive changes. Dividing this by the sum across all sampled households of each household's average household size across baseline and follow-up (7.0 + 6.0 = 13.0) gives $\hat{p}_{down}$ = 0 ÷ 13 = 0.000.[67]

The household-size-weighted average of the number of years between interviews y is 3.07.

With sample size n = 2 interviewed households, population N of 1,000 households, confidence level of 70 percent (z = 1.04), and the α adjustment factor for this estimator (regardless of poverty line) of 1.14,[68] the margin of error ±c is about

$$\pm 0.090 \approx \pm \frac{1.04 \cdot 1.14}{3.07} \cdot \sqrt{\frac{0.057 \cdot (1-0.057) + 0 \cdot (1-0) + 2 \cdot 0.057 \cdot 0}{2}} \cdot \sqrt{\frac{1{,}000-2}{1{,}000-1}}.$$

The example's estimated net annual poverty-rate change is –1.9 percentage points (Figure 11), so the 70-percent margin of error is –1.9 – 9.0 = –10.9 to –1.9 + 0.9 = +7.1 percentage points. The estimate from this tiny sample of n = 2 is uninformative; the true net change could easily be negative, close to zero, or positive.

This example shows why margins of error are useful. Without them, program managers might believe that there was evidence that poverty rates fell by 1.9 percentage points per year even though the data in this sample is also consistent with widely different rates and directions of change.

---

[67] $\hat{p}_{down} - \hat{p}_{up}$ is the estimated net poverty-rate change. In this particular example, $\hat{p}_{down}$ happens to be zero, so $-\hat{p}_{up}$ equals the estimated net poverty-rate change.

[68] Schreiner, 2020.

A5.2.8 Margins of error for estimates of the annual net change in the number of poor people across two periods for one sample, scored twice

The lower (upper) limit of the margin of error for an estimate of annual net change in the number of poor people for one sample, scored twice is the average number of people in participating households from baseline to follow-up, multiplied by the lower (upper) limit of the margin of error of the estimated annual net change in the poverty rate.

To illustrate with the example in Section 3.3.4 for one sample scored twice, the estimated annual net change in the poverty rate is –1.9 percentage points. As just shown, the tiny sample size of $n = 2$ means that the 70-percent margin of error runs from –10.9 to +7.1 percentage points.

The estimated average number of on-going participating people is 5,600.[69] Thus, the lower limit of the 70-percent margin of error for the estimated annual net change in the number of poor people is 5,600 · (–0.109) ≈ –610 (a net decrease in poor people), and the upper limit is 5,600·(+0.093) ≈ +398 (a net increase in poor people).

A5.2.9 Margins of error for estimates of the annual net change in head-count poverty rates across two periods for two independent samples

The formula for the margin of error $\pm c$ is $\pm \frac{z \cdot \alpha}{y} \cdot \sqrt{\frac{2 \cdot \hat{p} \cdot (1-\hat{p})}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$,

where z, α, y, $\hat{p}$ and N are defined as above, and n is the sample size of interviewed households at both baseline and follow-up.

Illustrating with the example for two independent samples in Section 3:

- $z = 1.04$, assuming a desired confidence level is 70 percent
- $\alpha = 1.10$, the adjustment factor (regardless of poverty line) for this estimator[70]
- $y = 2.77$, the years between the average interview at baseline and follow-up
- $\hat{p} = 0.902$, the (unadjusted) estimate of the poverty rate at baseline
- $N = 850$, the average number of households across baseline (1,000) and follow-up (700)
- $n = 2$, the sample size in both baseline and follow-up

[69] The formula for margin of error for the estimated number of poor people ignores that the estimated number of people in participating households has its own margin of error. This understates the margin of error.
[70] Schreiner, 2020.

The margin of error $\pm c$ is $\pm 0.123 \approx \pm \frac{1.04 \cdot 1.10}{2.77} \cdot \sqrt{\frac{2 \cdot 0.902 \cdot (1 - 0.902)}{2}} \cdot \sqrt{\frac{850 - 2}{850 - 1}}$.

The example's estimated net annual poverty-rate change is –3.1 percentage points (Figure 12). Thus, the 70-percent margin of error is –3.1 – 12.3 = –15.4 percentage points to –3.1 + 12.3 = +9.2 percentage points. The tiny sample is again consistent with a true value in the population that is negative, close to zero, or positive. This shows why margins of error matter.

A5.2.10 Margins of error for estimates of the annual net change in the number of poor people across two periods for two independent samples

The lower (upper) limit of the margin of error for an estimate of annual net change in the number of poor people for two independent samples is the average number of people in participating households from baseline to follow-up, multiplied by the lower (upper) limit of the margin of error of the estimated annual net change in the poverty rate.

To illustrate, the example in Section 3 for two independent samples estimates the annual net change in the poverty rate as –3.1 percentage points. As just shown, the 70-percent margin of error runs from –15.4 to +9.2 percentage points.

The estimated average number of on-going participating people is 5,250.[71] Thus, the lower limit of the 70-percent margin of error for the estimated annual net change in the number of poor people per year is 5,250 · (–0.154) ≈ –809 (a net decrease in poor people), and the upper limit is 5,250 · (+0.092) ≈ +483 (a net increase in poor people).

[71] The formula for margin of error for the estimated number of poor people ignores that the estimated number of people in participating households has its own margin of error. This understates the margin of error.

Annex 6 Formulas for sample size

Before drawing a sample of households to interview, the formulas here can be used to calculate the sample size that corresponds to a program's:

- Desired margin of error for the eventual scorecard estimate, and
- Desired confidence level for the margin of error, and
- Pre-estimation guess of the true population value to be estimated

These formulas may or may not be useful, for several reasons.

First, programs often collect scorecard data but then fail to report and analyze it. In such cases, the entire project is a waste, so there is no point in worrying about sample size. This is why programs must plan and budget for reporting and analysis. If the remaining budget will not cover at least 1,000 interviews, then ignore the formulas below and do as many interviews as the budget allows.

Second, both psychological sample size and statistical sample size matter. On the one hand, samples smaller than $n = 300$ often seem too small. On the other hand, samples of at least $n = 1{,}000$ usually seem large enough.

Third, calculating an optimal sample size makes sense only if a program:

- Has reason to desire a particular margin of error or level of confidence[72]
- Plans to report and analyze margins of error

If margins of error are not understood or will not be reported and analyzed, then just interview as many participating households as the budget allows.

Fourth, sample-size calculations are sometimes unneeded. For example, using the scorecard for segmenting requires interviewing all relevant participants. Likewise, doing a basic check on the fulfillment of a pro-poor mission may be less costly if all in-coming participants are scored as a routine step of the in-take process rather than repeatedly deciding at the moment whether to score a given enrollee.

[72] Academic conventions for levels of confidence, applied to business, often imply unnecessarily large samples.

In sum, go ahead with the formulas below if you:

- Reserve resources for reporting and analysis
- Understand margins of error and will report and analyze them
- Plan to estimate net changes in poverty over time, and
- Have enough budget for at least 1,000 interviews at both baseline and follow-up

Otherwise:

- If checking a pro-poor mission, then score all in-coming participants at in-take
- If segmenting by poverty, then score all relevant participants
- If estimating changes in poverty, then score as many participants as the budget allows

A6.1 Sample-size formula for snapshot estimates of head-count-poverty rates in a single time period

In this case, the formula for the sample size n (the number of participating households to be interviewed) is $n = N \cdot \left(\frac{z^2 \cdot \alpha^2 \cdot \tilde{p} \cdot (1-\tilde{p})}{z^2 \cdot \alpha^2 \cdot \tilde{p} \cdot (1-\tilde{p}) + c^2 \cdot (N-1)} \right)$,

where n, c, z, α, and N are defined as in Annex 5, and $\tilde{p}$ is a before-estimation guess for the poverty rate to be estimated.[73]

The illustration below of the calculation of the sample size n uses these values:

- The population of participating households is N = 10,000
- The desired confidence level for the margin of error is 80 percent, so z = 1.28
- The poverty line is 150% of the national line, so α = 1.14 (Figure 2)
- The pre-estimation expected poverty rate is the all-Tanzania rate for 150% of the national line in 2018, so $\tilde{p}$ = 54.0 percent = 0.540 (Figure 1)
- The desired margin of error ±c = ±3.0 percentage points = ±0.030

Given these hypothetical values,

$$n = 10{,}000 \cdot \left(\frac{1.28^2 \cdot 1.14^2 \cdot 0.540 \cdot (1-0.540)}{1.28^2 \cdot 1.14^2 \cdot 0.540 \cdot (1-0.540) + 0.03^2 \cdot (10{,}000-1)} \right) \approx 556.$$

[73] If the population N is "large" relative to the expected sample size n, then the formula can be taken as $n = \left(\frac{\alpha \cdot z}{c} \right)^2 \cdot \tilde{p} \cdot (1-\tilde{p})$.

A6.2 Sample-size formula for estimates of annual net changes in head-count-poverty rates across two time periods with one sample scored twice

The formula for the number of households to interview at both baseline and follow-up n is:[74]

$$2 \cdot \left(\frac{z \cdot \alpha}{c} \right)^2 \cdot [-0.01 + 0.016 \cdot y + 0.56 \cdot p_{\text{pre-baseline}} \cdot (1 - p_{\text{pre-baseline}})] \cdot \sqrt{\frac{N-n}{N-1}},$$

where n, α, z, c, and N are defined as above, y is the number of years between baseline and follow-up, and $p_{\text{pre-baseline}}$ is the population's expected head-count poverty rate prior to the baseline interviews.

The illustration below for this formula uses the following values:

- The poverty line is 150% of the national line
- The desired confidence level for the margin of error is 80 percent, so $z = 1.28$
- $\alpha = 1.14$ (regardless of the scorecard or poverty line[75]
- The desired margin of error $\pm c = \pm 3.0$ percentage points = ± 0.030
- The number of years between baseline and follow-up is $y = 3$
- The pre-estimation expected pre-baseline poverty rate is the all-Tanzania rate for 150% of the national line: $p_{\text{pre-baseline}} = 54.0$ percent = 0.540 (Figure 1)
- The population of participating households is $N = 10{,}000$

Assuming N is large relative to n so that $\sqrt{\frac{N-n}{N-1}} \approx 1$, the baseline sample size n is

$$2 \cdot \left(\frac{1.28 \cdot 1.14}{0.03} \right)^2 \cdot [-0.01 + 0.016 \cdot 3 + 0.56 \cdot 0.540 \cdot (1 - 0.540)] \cdot 1 \approx 838.$$

The follow-up sample size is also 838.

[74] Schreiner, 2020.
[75] Schreiner, 2020.

A6.3 Sample-size formula for estimates of annual net changes in head-count-poverty rates across two time periods with two independent samples

This formula is two (2), multiplied by the formula for sample size for a snapshot estimate at a point in time. If n and $\tilde{p}$ are the same at both baseline and follow-up, then $n = 2 \cdot N \cdot \left(\frac{z^2 \cdot \alpha^2 \cdot \tilde{p} \cdot (1-\tilde{p})}{z^2 \cdot \alpha^2 \cdot \tilde{p} \cdot (1-\tilde{p}) + c^2 \cdot (N-1)} \right)$.[76]

There are n interviews at baseline, and n interviews at follow-up. For this estimator and regardless of the scorecard or poverty line, α = 1.10.[77]

To illustrate with the same hypothetical values as in the example just above (except that α = 1.10), the sample size at baseline n is:

$$2 \cdot 10{,}000 \cdot \left(\frac{1.28^2 \cdot 1.10^2 \cdot 0.540 \cdot (1-0.540)}{1.28^2 \cdot 1.10^2 \cdot 0.540 \cdot (1-0.540) + 0.03^2 \cdot (10{,}000-1)} \right) \approx 1{,}038.$$

The sample size at follow-up is also n = 1,038.

[76] If the N is large relative to n, then the formula is about $n = 2 \cdot \left(\frac{\alpha \cdot z}{c} \right)^2 \cdot \tilde{p} \cdot (1-\tilde{p})$.

[77] Schreiner, 2020.

References

Adams, Niall M.; and David J. Hand. (2000) "Improving the Practice of Classifier Performance Assessment", *Neural Computation*, Vol. 12, pp. 305–311, request.

Brown, Caitlin; Ravallion, Martin; and Dominique van de Walle. (2018) "A Poor Means Test? Econometric Targeting in Africa", *Journal of Development Economics*, Vol. 134, pp. 109–124, link.

Caire, Dean. (2004) "Building Credit Scorecards for Small-Business Lending in Developing Markets", link.

_____; and Mark Schreiner. (2012) "Cross-Tab Weighting for Credit Scorecards in Developing Markets", link.

Coady, David; Grosh, Margaret; and John Hoddinott. (2004) *Targeting of Transfers in Developing Countries*, link.

Dawes, Robyn M. (1979) "The Robust Beauty of Improper Linear Models in Decision Making", *American Psychologist*, Vol. 34, No. 7, pp. 571–582, link.

Diamond, Alexis; Gill, Michael; Rebolledo Dellepiane, Miguel Angel; Skoufias, Emmanuel; Vinha, Katja; and Yiqing Xu. (2016) "Estimating Poverty Rates in Target Populations: An Assessment of the Simple Poverty Scorecard and Alternative Approaches", World Bank Policy Research Working Paper No. 7793, link.

Dupriez, Olivier. (2018) "An Empirical Comparison of Machine-Learning Classification Algorithms", link.

Ferreira, Francisco H.G.; Chen, Shaohua; Dabalen, Andrew L.; Dikhanov, Yuri; Hamadeh, Nada; Jolliffe, Dean; Narayan, Ambar; Prydz, Espen Beer; Revenga, Ana; Sangraula, Prem; Serajuddin, Umar; and Nobuo Yoshida. (2016) "A Global Count of the Extreme Poor in 2012: Data Issues, Methodology, and Initial Results", *Journal of Economic Inequality*, Vol. 14, pp. 141–172, request.

Grosh, Margaret; and Judy L. Baker. (1995) "Proxy-Means Tests for Targeting Social Programs: Simulations and Speculation", World Bank LSMS Working Paper No. 118, link.

Hand, David J. (2006) "Classifier Technology and the Illusion of Progress", *Statistical Science*, Vol. 22, No. 1, pp. 1–15, link.

Haslett, Stephen. (2012) “Practical Guidelines for the Design and Analysis of Sample Surveys for Small-Area Estimation”, *Journal of the Indian Society of Agricultural Statistics*, Vol. 66, No. 1, pp. 203–212, request.

Hoadley, Bruce; and Robert M. Oliver. (1998) “Business Measures of Scorecard Benefit”, *IMA Journal of Mathematics Applied in Business and Industry*, Vol. 9, pp. 55–64, request.

IRIS Center. (2007) “Manual for the Implementation of USAID Poverty-Assessment Tools”, link.

Jolliffe, Dean; and Espen Beer Prydz. (2016) “Estimating International Poverty Lines from Comparable National Thresholds”, *Journal of Economic Inequality*, Vol. 14, pp. 185–198, request.

Kolesar, Peter; and Janet L. Showers. (1985) “A Robust Credit-Screening Model Using Categorical Data”, *Management Science*, Vol. 31, No. 2, pp. 124–133, link.

Lovie, Alexander D.; and Patricia Lovie. (1986) “The Flat-Maximum Effect and Linear Scoring Models for Prediction”, *Journal of Forecasting*, Vol. 5, pp. 159–168, request.

Matul, Michal; and Sean Kline. (2003) “Scoring Change: Prizma’s Approach to Assessing Poverty”, MFC Spotlight Note No. 4, link.

Myers, James H.; and Edward W. Forgy. (1963) “The Development of Numerical Credit-Evaluation Systems”, *Journal of the American Statistical Association*, Vol. 58, No. 303, pp. 779–806, link.

Narayan, Ambar; and Nobuo Yoshida. (2005) “Proxy-Means Tests for Targeting Welfare Benefits in Sri Lanka”, World Bank Report No. SASPR–7, link.

National Bureau of Statistics. (2019) “Household Budget Survey 2017/18: Key Indicators Report”, link.

_____. (2014) *Household Budget Survey Main Report 2011/12 (Mainland)*, link.

Navajas, Sergio; Schreiner, Mark; Meyer, Richard L.; Gonzalez-Vega, Claudio; and Jorge Rodriguez-Meza. (2000) “Microfinance and the Poorest of the Poor: Theory and Evidence from Bolivia”, *World Development*, Vol. 28, No. 2, pp. 333–346, link.

Ravallion, Martin. (2009) “How Relevant Is Targeting to the Success of an Anti-Poverty Program?”, *World Bank Research Observer*, Vol. 24, No. 2, pp. 205–231, request.

_____. (1998) “Poverty Lines in Theory and Practice”, World Bank LSMS Working Paper No. 133, link.

Rutstein, Shea Oscar; and Kiersten Johnson. (2004) “The DHS Wealth Index”, DHS Comparative Reports No. 6, link.

Schreiner, Mark. (2020) “Scorocs Simple Poverty Scorecard Tool: Uganda”, request.

_____. (2016a) “Scorocs Simple Poverty Scorecard Tool: Tanzania”, link.

_____. (2016b) “Popular Science”, link.

_____. (2016c) “Simple Poverty Scorecard Tool: Indonesia (Jawa Timur and Nusa Tengarra Timur)”, request.

_____. (2015) “There’s No Place Like Home? How the Interview Method Affects Results with the Progress out of Poverty Index®”, link.

_____. (2014) “The Process of Poverty-Scoring Analysis”, link.

_____. (2013a) “Scorocs Simple Poverty Scorecard Tool: Tanzania”, link.

_____. (2013b) “Scorocs Simple Poverty Scorecard Tool: Bangladesh”, link.

_____. (2012) “An Expert-Based Poverty Scorecard for Rural China”, link.

_____. (2006) “Is One Simple Poverty Scorecard Enough for India?”, link.

_____. (2005) “La Herramienta del Índice de Calificación de la Pobreza™ de Scorocs: México”, link.

_____. (2002) *Scoring: The Next Breakthrough in Microfinance?* CGAP Occasional Paper No. 7, link.

_____. (1997) *A Framework for the Analysis of the Performance and Sustainability of Subsidized Microfinance Organizations with Application to BancoSol of Bolivia and the Grameen Bank of Bangladesh*, link.

_____; Matul, Michal; Pawlak, Ewa; and Sean Kline. (2014) “Poverty Scoring: Lessons from a Microlender in Bosnia-Herzegovina”, *Poverty and Public Policy*, Vol. 6, No. 4, pp. 407–428, link.

Sharif, Iffath Anwar. (2009) “Building a Targeting System for Bangladesh Based on Proxy-Means Testing”, World Bank Social Protection Discussion Paper No. 0914, link.

Stillwell, William G.; Barron, F. Hutton; and Ward Edwards. (1983) "Evaluating Credit Applications: A Validation of Multi-Attribute Utility-Weight Elicitation Techniques", *Organizational Behavior and Human Performance*, Vol. 32, pp. 87–108, request.

Tarozzi, Alessandro; and Angus Deaton. (2009) "Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas", *Review of Economics and Statistics*, Vol. 91, No. 4, pp. 773–792, link.

Toohig, Jeff. (2008) "PPI® Pilot Training Guide", link.

Wainer, Howard. (1976) "Estimating Coefficients in Linear Models: It Don't Make No Nevermind", *Psychological Bulletin*, Vol. 83, pp. 223–227, link.

World Bank. (2020) *Purchasing Power Parities and the Size of World Economies: Results from the 2017 International Comparison Program*, link.

_____. (2015) *Tanzania Mainland Poverty Assessment*, link.

_____. (2013) "Shared Prosperity: A New Goal for a Changing World", link.

_____. (2012) *Targeting Poor and Vulnerable Households in Indonesia*, link.

Zeller, Manfred. (2004) "Review of Poverty Assessment Tools", link.