Simple Poverty Scorecard® Tool
Ukraine

Mark Schreiner

1 September 2021

Таблиця та додаток для збору даних українською мовою знаходяться на 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 Ukraine 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

Acknowledgements
Chalice funded this paper, and the State Statistics Committee of Ukraine provided the data. Thanks go to Alisha Lebouthillier, Mikhail Matytsin, MaryAnne McKinnon, Volodymyr Sarioglo, and especially Lyubov Tsytsyk.

Scorocs® Simple Poverty Scorecard® Tool

			Name	Identifier
Interview ID:				
Interview date:		Participant of record:		
Country:	UKR	Field agent:		
Scorecard:	001	Service point:		
Sampling weight:		Number of household members:		

Question	Response	Points
1. In which region does the household live? (*From enumerator knowledge*)	A. Kharkiv, or Ivano-Frankivsk	0
	B. Cherkasy, Volyn, or Sumy	1
	C. City of Kyiv, Kyiv (excluding the city of Kyiv), Kherson, Rivne, or Chernivtsi	2
	D. Zakarpattya, or Khmelnytskiy	4
	E. Odesa, Zaporizhzhya, Vinnytsya, Poltava, Mykolayiv, or Luhansk	5
	F. Donetsk, Zhytomyr, or Kirovohrad	6
	G. Lviv, Dnipropetrovsk, Chernihiv, or Ternopil	10
2. How many members does the household have? (*From "Back-page Worksheet"*)	A. Four or more	0
	B. Three	12
	C. Two	20
	D. One	24
3. Is the head of the household male? (*From "Back-page Worksheet"*)	A. No	0
	B. Yes	6
4. How many members of the household are wage/salary employees? (*From "Back-page Worksheet"*)	A. None	0
	B. One	12
	C. Two or more	25
5. How many rooms does the household's residence have?	A. One	0
	B. Two	5
	C. Three or more	7
6. Does the household's residence have hot water?	A. No	0
	B. Yes	6
7. Does the household's residence have sewer service?	A. No	0
	B. Yes	5
8. Does the household have a clothes-washing machine?	A. No	0
	B. Yes	7
9. Does the household have a microwave oven?	A. No	0
	B. Yes	5
10. In the past three months, did the household keep any farm animals of its own, such as livestock, poultry, bees, and so on?	A. No	0
	B. Yes	5

  Score:

Back-page Worksheet: Members of the Household, and Work Status

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 participant of record (who may differ from the respondent), of the field agent of the participant of record (who may differ from you the enumerator), and of the service point that the participant of record uses (if any and if known). Circle the response to the first scorecard question ("In which region does the household live?") based on what you know, without asking the respondent.

Then read to the respondent: *Please tell me the first name (or nickname) and age of each household member, starting with the head. A* household *is one person alone or a group of people—regardless of blood or marital relationships—who live together in a residence, who keep house together, and who partly or fully share funds. Members of the household include people who live permanently with the household, even if they are temporarily absent on the day of the interview (as long as their total expected absence is 12 months or less).*

Write down the name and age of each member, first for the head and then for his/her spouse (if there is one). Record the sex of the head and of his/her spouse (if there is one). For each member 6-years-old or older, ask: "Does [NAME] work as a wage/salary employee?", and record the response.

After you finish with all household members, record the number of 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. For the third scorecard question, mark whether the head of the household is male. Finally, count the number of members who work as wage/salary employees, and record the response to the third scorecard question.

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

First name or nickname?	Head or spouse of head?	Is [NAME] 6-years-old or older?	If [Name] is at least 6-years-old, then ask: "Does [NAME] work as a wage/salary employee?"
1.	Head (male) Head (female)	No Yes	<6 years No Yes
2.	Spouse (female) Spouse (male) Other	No Yes	<6 years No Yes
3.	Other	No Yes	<6 years No Yes
4.	Other	No Yes	<6 years No Yes
5.	Other	No Yes	<6 years No Yes
6.	Other	No Yes	<6 years No Yes
7.	Other	No Yes	<6 years No Yes
8.	Other	No Yes	<6 years No Yes
9.	Other	No Yes	<6 years No Yes
10.	Other	No Yes	<6 years No Yes
11.	Other	No Yes	<6 years No Yes
12.	Other	No Yes	<6 years No Yes
13.	Other	No Yes	<6 years No Yes
Number of household members:			Number wage/salary employees:

Figure 1: Conversion of scores to poverty likelihoods

	Poverty likelihood (%)									
	National			Percentile-based lines						
Score	100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
0–34	71.8	94.2	99.3	43.2	60.4	74.4	84.9	91.1	98.9	99.0
35–39	58.2	90.7	98.8	20.2	40.8	65.2	76.4	83.9	94.5	98.4
40–42	54.8	86.8	97.5	16.1	32.6	59.5	69.0	75.1	90.6	97.8
43–44	40.6	86.5	97.3	15.4	25.1	46.8	58.9	70.4	90.1	97.0
45–46	40.6	86.5	97.3	15.4	23.8	46.8	58.9	70.4	90.1	97.0
47–48	40.6	83.5	96.6	15.2	23.2	45.3	58.5	70.4	89.7	96.6
49–50	35.1	80.3	93.9	6.9	22.2	38.9	56.0	70.4	88.1	94.1
51–52	26.6	77.9	92.0	6.6	16.5	37.0	47.7	62.0	83.5	92.3
53–54	26.6	73.5	90.7	6.6	16.5	37.0	47.7	62.0	82.8	92.1
55–56	26.6	73.5	90.7	5.3	16.3	36.6	47.7	62.0	82.8	92.1
57–58	19.6	64.5	87.2	4.4	10.0	28.8	40.6	52.5	72.7	88.7
59–60	14.4	56.8	81.7	2.3	9.2	25.6	35.2	45.9	69.5	84.9
61–62	11.7	53.3	80.6	2.3	5.8	22.1	33.4	44.3	69.5	84.9
63–64	8.3	50.6	80.6	1.7	4.5	17.9	30.7	42.4	69.5	84.9
65–66	8.3	50.6	80.6	1.2	3.9	17.9	29.7	42.4	69.5	84.9
67–69	6.0	36.9	67.1	1.2	3.9	12.0	16.2	23.0	57.9	74.2
70–72	5.8	32.5	61.6	0.8	3.9	10.6	15.2	21.6	52.1	68.8
73–77	5.8	30.6	58.8	0.5	3.9	9.4	14.2	18.0	42.0	68.4
78–100	3.1	21.1	51.4	0.2	2.1	4.8	9.4	12.3	30.6	58.5

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			Percentile-based lines						
	100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
Estimation error	+1.6	–4.3	–3.5	–2.6	–4.3	–13.1	–11.0	–11.5	–7.3	–4.4
Margin of error	2.6	3.0	2.0	2.1	2.7	3.3	3.3	3.1	2.4	1.7
α factor	1.24	1.18	0.79	1.57	1.50	1.46	1.35	1.22	1.03	0.96

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
Ukraine

1. Introduction

The Scorocs Simple Poverty Scorecard-brand poverty-assessment tool for Ukraine 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 income and expenditure surveys. A case in point is the 2018 Household Living Conditions Survey (HLCS, Вибіркове Обстеження Умов Життя Домогосподарств) by the State Statistics Service of Ukraine (SSSU). The 2018 HLCS has more than 20 pages and asks more than 200 top-level questions, many of which have several follow-up questions or are repeated (for example, for each household member or for each plot of land).

1.2 How the scorecard works

The scorecard has 10 factual questions that are drawn from the exhaustive 2018 HLCS. Examples include: “How many rooms does the household’s residence have?” and “Does the household have a microwave oven?”.

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 Ukraine

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 income-based poverty in the 2018 HLCS.

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.[1]

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 HLCS 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

---

[1] 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"[2]—the scorecard is transparent, freely available,[3] 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 Income-based poverty

Ukraine's scorecard is a quantitative way to assess whether a program's participants have income below any of 10 poverty lines. The most-relevant line is probably Ukraine's "actual subsistence minimum" line (called here the "100% of the national line") of about UAH107 per adult equivalent per day, giving a country-wide head-count poverty rate of 26.7 percent in 2018

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 a national line while internally using a 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.

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

[3] Ukraine's scorecard is not in the public domain; it is copyright © 2021 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.[4] 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.[5] 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.[6] In particular:

- A given program's participants are not representative of Ukraine 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 HLCS. 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)[7]

---

[4] Schreiner, 2002.

[5] 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.

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

[7] For example, the 2020/21 economic downturn due to COVID–19 changed the links between poverty and questions, but the Ukraine 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 HLCS 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 to 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 Ukraine's 2018 HLCS 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 ("UKR", pre-filled)
  - Scorecard code ("001", 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:
  - *Participant of record*. This is the member of the household whose identifying information is recorded with the pro-poor program. Often, the participant of record is the adult member of the household who interacts directly with the program. He/she may or may not be the same as the respondent who responds to the scorecard questions. For example, a participant of record for a microfinance program is a borrower or a saver, and a participant of record with a child-health program might be a child or a child's parent or guardian
  - *Field agent* (if there is one). This is the participant of record'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 participant of record. The service point is usually the base of operations ofor the field agent who serves the participant of record (if there is one) or where the participant of record 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 almost always 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), age, and work status
- 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?")
  - — The response to the third scorecard question ("Is the head of the household male?")
  - — The response to the fourth scorecard question ("How many members of the household are wage/salary employees?")
- Read the remaining six 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 the participant of record, 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, 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 Ukraine's SSSU in the 2018 HLCS. 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 data.

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.[8] 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.[9] 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 that there is an incentive to exaggerate poverty estimates (for example, if they expect to be rewarded for higher poverty rates).[10]

---

[8] 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 Ukraine's SSSU did in the 2018 HLCS.

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

[10] 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

Interview ID:	A123		**Name**	**Identifier**
Interview date:	13JUN2021	Participant of record:	ANNA JACKSON	1V0276FZ7
Country:	UKR	Field agent:	UNKNOWN	UNKNOWN
Scorecard:	001	Service point:	NORTHWEST CLINIC	NWC
Sampling weight:	UNKNOWN	Number of household members:		FIVE

Question	Response	Points	
1. In which region does the household live? (*From enumerator knowledge*)	A. Kharkiv, or Ivano-Frankivsk	0	
	B. Cherkasy, Volyn, or Sumy	1	
	C. City of Kyiv, Kyiv (excluding the city of Kyiv). Kherson, Rivne, or Chernivtsi	2	
	D. Zakarpattya, or Khmelnytskiy	4	
	E. Odesa, Zaporizhzhya, Vinnytsya, Poltava, Mykolayiv, or Luhansk	5	
	F. Donetsk, Zhytomyr, or Kirovohrad	6	
	G. Lviv, Dnipropetrovsk, Chernihiv, or Ternopil	10	10
2. How many members does the household have? (*From "Back-page Worksheet"*)	A. Four or more	0	0
	B. Three	12	
	C. Two	20	
	D. One	24	
3. Is the head of the household male? (*From "Back-page Worksheet"*)	A. No	0	0
	B. Yes	6	
4. How many members of the household are wage/salary employees? (*From "Back-page Worksheet"*)	A. None	0	
	B. One	12	12
	C. Two or more	25	
5. How many rooms does the household's residence have?	A. One	0	
	B. Two	5	
	C. Three or more	7	7
6. Does the household's residence have hot water?	A. No	0	
	B. Yes	6	6
7. Does the household's residence have sewer service?	A. No	0	0
	B. Yes	5	
8. Does the household have a clothes-washing machine?	A. No	0	0
	B. Yes	7	
9. Does the household have a microwave oven?	A. No	0	0
	B. Yes	5	
10. In the past three months, did the household keep any farm animals of its own, such as livestock, poultry, bees, and so on?	A. No	0	0
	B. Yes	5	

scorocs.com

Score: 10 + 0 + 0 + 12 + 7 + 6 + 0 + 0 + 0 + 0 = 35

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

First name or nickname?	Head or spouse of head?	Is [NAME] 6-years-old or older?		If [Name] is at least 6-years-old, then ask: "Does [NAME] work as a wage/salary employee?"		
1. ANNA	Head (male) Head (female)	No	Yes	<6 years	No	Yes
2. BILLY	Spouse (female) Spouse (male) Other	No	Yes	<6 years	No	Yes
3. CHARLES	Other	No	Yes	<6 years	No	Yes
4. DARLA	Other	No	Yes	<6 years	No	Yes
5. EUGENE	Other	No	Yes	<6 years	No	Yes
6.	Other	No	Yes	<6 years	No	Yes
. . .	Other	No	Yes	<6 years	No	Yes
13.	Other	No	Yes	<6 years	No	Yes
Number of household members: FIVE				Number wage/salary employees: ONE		

2.3 First example household

The points for the first example household's responses add up to a score of 35 (Figure 3 and Figure 4).

For a given poverty line, Figure 1 lists poverty likelihoods by score range. A score of 35 falls in the second range of 35–39. For 100% of the national poverty line, the poverty likelihood for scores of 35–39 is 58.2 percent. That is, the scorecard estimates that 58.2 percent of households in Ukraine with a score of 35–39 have income below 100% of the national line.

Figure 5: The first example household's score of 35 implies a poverty likelihood of 58.2 percent for 100% of the national line (excerpted from Figure 1)

	Poverty likelihood (%)		
		National	
Score	100%	150%	200%
0–34	71.8	94.2	99.3
35–39	58.2	90.7	98.8
40–42	54.8	86.8	97.5
43–44	40.6	86.5	97.3
45–46	40.6	86.5	97.3
47–48	40.6	83.5	96.6
49–50	35.1	80.3	93.9
51–52	26.6	77.9	92.0
53–54	26.6	73.5	90.
. . .	. . .	. . .	. . .

Figure 6: Second example household, filled-in scorecard

			Name	Identifier
Interview ID:	B456			
Interview date:	30JUN2021	Participant of record:	JOHN BROWN	2W3120ZG8
Country:	UKR	Field agent:	UNKNOWN	UNKNOWN
Scorecard:	001	Service point:	NORTHWEST CLINIC	NWC
Sampling weight:	UNKNOWN	Number of household members:		FOUR

Question	Response	Points	
1. In which region does the household live? (*From enumerator knowledge*)	A. Kharkiv, or Ivano-Frankivsk	0	
	B. Cherkasy, Volyn, or Sumy	1	
	C. City of Kyiv, Kyiv (excluding the city of Kyiv), Kherson, Rivne, or Chernivtsi	2	
	D. Zakarpattya, or Khmelnytskiy	4	
	E. Odesa, Zaporizhzhya, Vinnytsya, Poltava, Mykolayiv, or Luhansk	5	
	F. Donetsk, Zhytomyr, or Kirovohrad	6	
	G. Lviv, Dnipropetrovsk, Chernihiv, or Ternopil	10	10
2. How many members does the household have? (*From "Back-page Worksheet"*)	A. Four or more	0	0
	B. Three	12	
	C. Two	20	
	D. One	24	
3. Is the head of the household male? (*From "Back-page Worksheet"*)	A. No	0	
	B. Yes	6	6
4. How many members of the household are wage/salary employees? (*From "Back-page Worksheet"*)	A. None	0	
	B. One	12	12
	C. Two or more	25	
5. How many rooms does the household's residence have?	A. One	0	
	B. Two	5	5
	C. Three or more	7	
6. Does the household's residence have hot water?	A. No	0	0
	B. Yes	6	
7. Does the household's residence have sewer service?	A. No	0	0
	B. Yes	5	
8. Does the household have a clothes-washing machine?	A. No	0	0
	B. Yes	7	
9. Does the household have a microwave oven?	A. No	0	
	B. Yes	5	5
10. In the past three months, did the household keep any farm animals of its own, such as livestock, poultry, bees, and so on?	A. No	0	
	B. Yes	5	5

Score: 10 + 0 + 6 + 12 + 5 + 0 + 0 + 0 + 5 + 5 = 43

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

First name or nickname?	Head or spouse of head?	Is [NAME] 6-years-old or older?	If [Name] is at least 6-years-old, then ask: “Does [NAME] work as a wage/salary employee?”
1. JOHN	Head (male) Head (female)	No Yes	<6 years No Yes
2. MARY	Spouse (female) Spouse (male) Other	No Yes	<6 years No Yes
3. SUE	Other	No Yes	<6 years No Yes
4. KIM	Other	No Yes	<6 years No Yes
6.	Other	No Yes	<6 years No Yes
. . .	Other	No Yes	<6 years No Yes
13.	Other	No Yes	<6 years No Yes
Number of household members: FOUR			Number wage/salary employees: ONE

2.4 Second example household

The points for the second example household's responses add up to a score of 43 (Figure 6 and Figure 7).

In Figure 1, a score of 43 falls in the range of 43–44. For 100% of the national poverty line, the poverty likelihood for scores of 43–44 is 40.6 percent. The scorecard estimates that 40.6 percent of households in Ukraine with a score of 43–44 have income below 100% of the national line.

Figure 8: The second example household's score of 43 implies a poverty likelihood of 40.6 percent for 100% of the national line (excerpt from Figure 1)

	Poverty likelihood (%)		
		National	
Score	100%	150%	200%
0–34	71.8	94.2	99.3
35–39	58.2	90.7	98.8
40–42	54.8	86.8	97.5
43–44	40.6	86.5	97.3
45–46	40.6	86.5	97.3
47–48	40.6	83.5	96.6
49–50	35.1	80.3	93.9
51–52	26.6	77.9	92.0
53–54	26.6	73.5	90.
. . .	. . .	. . .	. . .

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 income (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 urban Ternopil in 2021. In that calendar year, it enrolls 1,000 in-coming households, from which it scores a simple random sample[11] of two households.[12]

The program judges that 100% 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 +1.6 percentage points (Figure 2).

The first example household has five members and is interviewed on June 13, 2021 (Figure 3 and Figure 4). With a score of 35, it has a poverty likelihood for 100% of the national line of 58.2 percent (Figure 1).

The second example household has four members and is interviewed on June 30, 2021 (Figure 6 and Figure 7). Its score of 43 corresponds with a poverty likelihood of 40.6 percent.

---

[11] 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.

[12] 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 here.

The estimated head-count poverty rate for the population of in-coming households in the 2021 calendar-year cohort in this new urban Ternopil branch 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:

$$\frac{5 \cdot 0.582 + 4 \cdot 0.406}{5+4} - (+0.016) \approx \frac{4.53}{9} - 0.016 \approx 0.488 = 48.8 \text{ percent.}$$

The five in the "$5 \cdot 0.582$" term is the number of members (household size) in the first household, and 0.582 is the first household's estimated poverty likelihood as proportion.

In the same way, the four "$4 \cdot 0.406$" is the number of members in the second household, and 0.406 is the second household's estimated poverty likelihood.

The "5 + 4" is the sum of the weights—that is, the number of household members—across the two sampled households.

The "+0.016" is the scorecard's estimation error for this poverty line (Figure 2). Because unadjusted estimates tend to be too high by 1.6 percentage points, they are adjusted downwards by subtracting +1.6. This is akin to how an archer whose arrows tend to miss a little to the right of the bulls-eye will adjust his/her aim to be a little to the left of the bulls-eye.

The estimated head-count poverty rate for the population is 48.8 percent. Again, this is the household-size-weighted average of the two sampled households' poverty likelihoods, adjusted for the known estimation error.[13]

With hundreds or thousands of interviewed households, the calculations shoud be done by an app or in a spreadsheet modeled on Figure 9 below.

---

[13] Be careful; the estimated poverty rate is *not* the single poverty likelihood associated with the household-size-weighted average score, which here is (5·35 + 4·43) ÷ (5 + 4) ≈ 39. This average score of 39 corresponds to a poverty likelihood of 58.2 percent (Figure 1), giving an error-adjusted poverty rate of 58.2 – (+1.6) = 56.6 percent. This differs from the 48.8 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-21	1V0276FZ7	5	35	58.2	2.91 = (D2*F2)/100
3	Baseline	30-Jun-21	2W3120ZG8	4	43	40.6	1.62 = (D3*F3)/100
4			Sum:	9 = SUM(D2:D3)			4.53 = SUM(G2:G3)
5			Average:	4.5 = AVERAGE(D2:D3)			
6							
7	Estimated scorecard error for this poverty line (percentage points):						+1.6
8							
9				Estimated head-count poverty rate (%):			48.8 = (G4/D4)*100-G7
10							
11				Households in the population:			1,000
12							
13				People in households in the population:			4,500 = G11*D5
14							
15				Number of poor people in population:			2,195 = (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*.[14] 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 HLCS for all 10 poverty lines by urban/rural/all for Ukraine overall and for each of its 25 regions. In the example of urban Ternopil, the head-count poverty rate for 100% of the national line is 15.1 percent. Thus, the example program is pro-poor in the sense that its in-coming participants have an above-average estimated poverty rate (48.8 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 nine pages. The regions in **Figure 10** begin with all-Ukraine overall, followed by the 25 regions in the SSSU's traditional order for reporting.

---

[14] The Ukraine scorecard uses an income-based definition of *poverty*. Common non-income 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: (Ukraine overall, Vinnytsya, and Volyn): 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			Percentile-based lines						
				100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
Ukraine	Urban	Line	4,998	107	161	215	67	79	97	107	120	155	195
		Rate		24.7	66.2	85.1	9.1	17.8	38.8	49.1	58.8	78.3	88.1
	Rural	Line	3,031	107	161	215	67	79	97	107	120	155	195
		Rate		30.9	70.6	90.3	12.0	23.8	41.8	51.4	62.1	83.4	93.7
	All	Line	8,039	107	161	215	67	79	97	107	120	155	195
		Rate		26.7	67.7	86.8	10.0	20.0	40.0	50.0	60.0	80.0	90.0
Vinnytsya	Urban	Line	167	107	161	215	67	79	97	107	120	155	195
		Rate		20.6	61.1	83.9	6.9	15.9	30.1	40.1	52.2	77.9	88.6
	Rural	Line	154	107	161	215	67	79	97	107	120	155	195
		Rate		31.7	74.3	90.6	14.9	28.3	41.7	54.1	62.2	83.0	91.8
	All	Line	321	107	161	215	67	79	97	107	120	155	195
		Rate		26.0	67.6	87.2	10.8	21.9	35.8	46.9	57.1	80.4	90.2
Volyn	Urban	Line	129	107	161	215	67	79	97	107	120	155	195
		Rate		31.6	73.3	94.6	16.4	26.3	50.0	59.4	68.2	90.7	96.2
	Rural	Line	96	107	161	215	67	79	97	107	120	155	195
		Rate		44.9	71.6	85.9	8.8	39.2	48.8	55.3	66.4	80.6	94.7
	All	Line	225	107	161	215	67	79	97	107	120	155	195
		Rate		37.8	72.5	90.5	12.8	32.3	49.4	57.5	67.4	86.0	95.5

Source: 2018 HLCS.

Poverty rates are percentages.

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

Percentile-based lines are UAH per-person, per-day.

All poverty lines are UAH in prices in Ukraine as a whole on average during the 2018 HLCS fieldwork.

Figure 10: (Dnipropetrovsk, Donetsk, and Zhytomyr): Poverty lines and head-count poverty rates by urban/rural/all in 2018

Region/ Area		Line or Rate	*n*	National 100%	National 150%	National 200%	Percentile-based lines 10th	20th	40th	50th	60th	80th	90th
Dnipropetrovsk	Urban	Line	287	107	161	215	67	79	97	107	120	155	195
		Rate		15.3	58.3	82.4	6.4	9.9	22.5	38.0	48.1	68.2	88.1
	Rural	Line	131	107	161	215	67	79	97	107	120	155	195
		Rate		35.6	68.5	83.3	19.5	31.9	39.1	48.4	59.1	76.9	87.9
	All	Line	418	107	161	215	67	79	97	107	120	155	195
		Rate		18.6	60.0	82.5	8.6	13.5	25.2	39.7	49.9	69.6	88.0
Donetsk	Urban	Line	426	107	161	215	67	79	97	107	120	155	195
		Rate		30.3	74.2	85.5	12.0	18.7	38.0	53.2	60.3	81.6	88.1
	Rural	Line	117	107	161	215	67	79	97	107	120	155	195
		Rate		41.3	76.2	91.4	6.8	34.4	46.4	52.1	68.8	81.3	95.6
	All	Line	543	107	161	215	67	79	97	107	120	155	195
		Rate		32.1	74.6	86.4	11.2	21.3	39.4	53.1	61.6	81.5	89.3
Zhytomyr	Urban	Line	130	107	161	215	67	79	97	107	120	155	195
		Rate		31.6	74.5	91.5	5.3	16.4	51.9	63.1	68.7	84.5	91.5
	Rural	Line	119	107	161	215	67	79	97	107	120	155	195
		Rate		32.9	74.1	83.8	16.6	24.1	43.2	54.0	66.0	80.7	91.5
	All	Line	249	107	161	215	67	79	97	107	120	155	195
		Rate		32.1	74.3	88.4	9.9	19.6	48.4	59.4	67.6	82.9	91.5

Source: 2018 HLCS.

Poverty rates are percentages.

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

Percentile-based lines are UAH per-person, per-day.

All poverty lines are UAH in prices in Ukraine as a whole on average during the 2018 HLCS fieldwork.

Figure 10: (Zakarpattya, Zaporizhzhya, and Ivano-Frankivsk): 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			Percentile-based lines						
				100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
Zakarpattya	Urban	Line	98	107	161	215	67	79	97	107	120	155	195
		Rate		23.2	76.5	90.0	15.8	41.8	60.9	65.3	75.0	84.9	90.1
	Rural	Line	98	107	161	215	67	79	97	107	120	155	195
		Rate		39.5	71.4	87.9	18.8	31.3	48.3	58.8	70.0	83.2	94.7
	All	Line	198	107	161	215	67	79	97	107	120	155	195
		Rate		33.4	73.6	88.8	17.5	35.8	53.5	61.6	72.1	84.0	92.9
Zaporizhzhya	Urban	Line	246	107	161	215	67	79	97	107	120	155	195
		Rate		24.5	64.3	87.9	9.3	18.2	34.0	38.8	54.8	81.0	91.8
	Rural	Line	120	107	161	215	67	79	97	107	120	155	195
		Rate		24.6	86.4	95.4	7.1	18.7	43.0	54.5	76.5	89.6	95.7
	All	Line	366	107	161	215	67	79	97	107	120	155	195
		Rate		24.5	69.4	89.6	8.8	18.4	36.0	42.4	59.7	83.0	92.7
Ivano-Frankivsk	Urban	Line	149	107	161	215	67	79	97	107	120	155	195
		Rate		17.5	69.2	91.9	9.7	11.7	30.0	48.6	59.8	70.5	92.9
	Rural	Line	110	107	161	215	67	79	97	107	120	155	195
		Rate		27.3	62.1	88.2	6.4	24.9	34.0	44.0	54.1	86.2	99.5
	All	Line	259	107	161	215	67	79	97	107	120	155	195
		Rate		23.1	65.1	89.8	7.8	19.2	32.3	46.0	56.6	79.4	96.7

Source: 2018 HLCS.

Poverty rates are percentages.

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

Percentile-based lines are UAH per-person, per-day.

All poverty lines are UAH in prices in Ukraine as a whole on average during the 2018 HLCS fieldwork.

Figure 10: (Kyiv (not Kyiv city), Kirovohrad, and Luhansk): 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			Percentile-based lines						
				100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
Kyiv (not Kyiv city)	Urban	Line	163	107	161	215	67	79	97	107	120	155	195
		Rate		34.0	73.4	88.9	13.6	26.5	44.2	54.5	66.7	84.7	89.2
	Rural	Line	106	107	161	215	67	79	97	107	120	155	195
		Rate		34.4	84.3	95.4	16.0	23.0	48.4	62.5	73.4	88.3	94.6
	All	Line	269	107	161	215	67	79	97	107	120	155	195
		Rate		34.2	77.6	91.4	14.5	25.2	45.8	57.6	69.3	86.1	91.3
Kirovohrad	Urban	Line	154	107	161	215	67	79	97	107	120	155	195
		Rate		18.7	66.1	85.3	4.9	12.3	30.4	39.3	55.3	77.0	91.9
	Rural	Line	107	107	161	215	67	79	97	107	120	155	195
		Rate		30.4	76.6	87.9	18.4	26.5	42.9	54.2	66.3	83.9	88.7
	All	Line	261	107	161	215	67	79	97	107	120	155	195
		Rate		23.1	70.0	86.3	10.0	17.7	35.1	44.9	59.5	79.6	90.7
Luhansk	Urban	Line	389	107	161	215	67	79	97	107	120	155	195
		Rate		28.3	76.3	87.6	11.5	20.7	49.5	57.8	64.7	81.4	88.9
	Rural	Line	125	107	161	215	67	79	97	107	120	155	195
		Rate		20.8	88.6	95.1	4.2	7.0	50.6	62.3	79.4	90.1	94.0
	All	Line	514	107	161	215	67	79	97	107	120	155	195
		Rate		25.9	80.1	90.0	9.2	16.4	49.8	59.2	69.3	84.1	90.5

Source: 2018 HLCS.

Poverty rates are percentages.

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

Percentile-based lines are UAH per-person, per-day.

All poverty lines are UAH in prices in Ukraine as a whole on average during the 2018 HLCS fieldwork.

Figure 10: (Lviv, Mykolayiv, and Odesa): 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			Percentile-based lines						
				100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
Lviv	Urban	Line	234	107	161	215	67	79	97	107	120	155	195
		Rate		19.6	62.7	82.3	5.7	15.4	34.4	49.8	59.7	78.1	86.3
	Rural	Line	184	107	161	215	67	79	97	107	120	155	195
		Rate		12.7	54.7	89.2	3.3	8.3	18.8	27.4	48.8	81.9	92.5
	All	Line	420	107	161	215	67	79	97	107	120	155	195
		Rate		16.3	61.2	85.5	4.6	16.0	31.2	43.5	57.3	80.3	89.0
Mykolayiv	Urban	Line	241	107	161	215	67	79	97	107	120	155	195
		Rate		27.8	80.2	94.2	8.8	26.8	56.0	59.5	69.1	89.1	94.7
	Rural	Line	133	107	161	215	67	79	97	107	120	155	195
		Rate		31.0	82.9	94.0	13.6	24.2	66.1	72.6	77.0	90.0	96.0
	All	Line	374	107	161	215	67	79	97	107	120	155	195
		Rate		28.7	81.0	94.2	10.3	26.0	59.1	63.5	71.5	89.4	95.1
Odesa	Urban	Line	183	107	161	215	67	79	97	107	120	155	195
		Rate		17.6	61.7	82.1	6.4	11.2	41.9	47.8	57.4	77.0	87.2
	Rural	Line	162	107	161	215	67	79	97	107	120	155	195
		Rate		24.9	68.4	88.6	6.9	18.2	45.0	54.9	63.5	79.7	96.9
	All	Line	347	107	161	215	67	79	97	107	120	155	195
		Rate		19.9	63.8	84.2	6.6	13.5	42.9	50.1	59.3	77.8	90.4

Source: 2018 HLCS.

Poverty rates are percentages.

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

Percentile-based lines are UAH per-person, per-day.

All poverty lines are UAH in prices in Ukraine as a whole on average during the 2018 HLCS fieldwork.

Figure 10: (Poltava, Rivne, and Sumy): 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			Percentile-based lines						
				100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
Poltava	Urban	Line	144	107	161	215	67	79	97	107	120	155	195
		Rate		19.8	57.9	75.9	7.4	14.1	33.1	46.3	52.5	69.8	82.5
	Rural	Line	139	107	161	215	67	79	97	107	120	155	195
		Rate		28.8	61.8	90.5	13.4	23.8	38.4	42.8	55.3	75.5	90.7
	All	Line	283	107	161	215	67	79	97	107	120	155	195
		Rate		23.3	59.4	81.5	9.7	17.8	35.2	45.0	53.6	72.0	85.7
Rivne	Urban	Line	147	107	161	215	67	79	97	107	120	155	195
		Rate		41.2	74.1	94.0	19.2	32.6	50.1	61.9	69.7	87.9	94.9
	Rural	Line	131	107	161	215	67	79	97	107	120	155	195
		Rate		34.9	73.8	94.7	10.8	21.2	45.8	54.6	64.8	90.1	97.0
	All	Line	278	107	161	215	67	79	97	107	120	155	195
		Rate		37.9	74.0	94.4	14.8	26.6	47.9	58.1	67.1	89.1	96.0
Sumy	Urban	Line	198	107	161	215	67	79	97	107	120	155	195
		Rate		43.7	72.8	91.3	17.3	29.6	53.1	58.3	66.6	86.9	94.3
	Rural	Line	136	107	161	215	67	79	97	107	120	155	195
		Rate		22.0	68.8	90.5	7.5	10.6	32.8	48.6	57.6	79.9	91.6
	All	Line	334	107	161	215	67	79	97	107	120	155	195
		Rate		37.0	71.5	91.1	14.3	23.7	46.8	55.3	63.8	84.7	93.4

Source: 2018 HLCS.

Poverty rates are percentages.

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

Percentile-based lines are UAH per-person, per-day.

All poverty lines are UAH in prices in Ukraine as a whole on average during the 2018 HLCS fieldwork.

Figure 10: (Ternopil, Kharkiv, and Kherson): Poverty lines and head-count poverty rates by urban/rural/all in 2018

Region/ Area		Line or Rate	n	National			Percentile-based lines						
				100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
Ternopil	Urban	Line	87	107	161	215	67	79	97	107	120	155	195
		Rate		15.1	81.7	94.5	1.6	6.7	48.0	59.1	79.8	92.5	97.5
	Rural	Line	131	107	161	215	67	79	97	107	120	155	195
		Rate		33.3	60.7	92.7	12.8	29.8	38.7	44.0	50.6	89.5	93.9
	All	Line	222	107	161	215	67	79	97	107	120	155	195
		Rate		25.4	70.7	93.6	7.6	19.9	43.8	51.6	64.4	91.0	95.6
Kharkiv	Urban	Line	212	107	161	215	67	79	97	107	120	155	195
		Rate		33.1	69.3	90.0	11.9	20.7	50.6	56.3	61.2	85.0	94.0
	Rural	Line	129	107	161	215	67	79	97	107	120	155	195
		Rate		26.1	75.1	93.6	6.8	13.3	40.7	56.1	65.2	84.9	93.5
	All	Line	341	107	161	215	67	79	97	107	120	155	195
		Rate		31.7	70.5	90.7	10.9	19.2	48.6	56.3	62.0	85.0	93.9
Kherson	Urban	Line	231	107	161	215	67	79	97	107	120	155	195
		Rate		37.0	80.9	96.3	14.4	25.9	52.1	67.0	75.1	92.0	95.9
	Rural	Line	162	107	161	215	67	79	97	107	120	155	195
		Rate		46.9	83.2	94.3	13.5	39.1	54.7	68.3	77.0	91.9	94.9
	All	Line	393	107	161	215	67	79	97	107	120	155	195
		Rate		40.9	81.8	95.5	14.0	31.2	53.1	67.5	75.9	92.0	95.5

Source: 2018 HLCS.

Poverty rates are percentages.

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

Percentile-based lines are UAH per-person, per-day.

All poverty lines are UAH in prices in Ukraine as a whole on average during the 2018 HLCS fieldwork.

Figure 10: (Khmelnytskiy, Cherkasy, and Chernivtsi): Poverty lines and head-count poverty rates by urban/rural/all in 2018

Region/ Area		Line or Rate	*n*	National 100%	National 150%	National 200%	Percentile-based lines 10th	20th	40th	50th	60th	80th	90th
Khmelnytskiy	Urban	Line	124	107	161	215	67	79	97	107	120	155	195
		Rate		34.3	79.0	92.2	6.2	31.2	56.5	65.6	74.6	90.0	93.0
	Rural	Line	117	107	161	215	67	79	97	107	120	155	195
		Rate		35.4	66.5	96.5	12.2	20.9	44.1	50.1	58.7	78.2	96.3
	All	Line	241	107	161	215	67	79	97	107	120	155	195
		Rate		34.8	73.5	94.1	8.9	26.7	51.0	58.7	67.6	84.8	94.4
Cherkasy	Urban	Line	256	107	161	215	67	79	97	107	120	155	195
		Rate		34.5	76.9	92.8	13.3	25.0	40.0	53.9	66.6	87.0	94.9
	Rural	Line	121	107	161	215	67	79	97	107	120	155	195
		Rate		37.8	78.0	93.0	19.6	32.1	49.8	58.3	60.9	86.4	93.5
	All	Line	377	107	161	215	67	79	97	107	120	155	195
		Rate		36.0	77.4	92.9	16.2	28.2	44.4	55.8	64.0	86.7	94.3
Chernivtsi	Urban	Line	125	107	161	215	67	79	97	107	120	155	195
		Rate		30.2	73.0	93.1	12.5	15.4	39.3	54.7	68.2	85.0	94.1
	Rural	Line	84	107	161	215	67	79	97	107	120	155	195
		Rate		24.8	55.2	91.6	17.3	20.4	33.3	36.1	43.5	86.2	91.1
	All	Line	209	107	161	215	67	79	97	107	120	155	195
		Rate		27.0	62.5	92.3	15.3	18.3	35.7	43.7	53.6	85.7	92.3

Source: 2018 HLCS.

Poverty rates are percentages.

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

Percentile-based lines are UAH per-person, per-day.

All poverty lines are UAH in prices in Ukraine as a whole on average during the 2018 HLCS fieldwork.

Figure 10: (Chernihiv, and City of Kyiv): Poverty lines and head-count poverty rates by urban/rural/all in 2018

Region/ Area		Line or Rate	*n*	National			Percentile-based lines						
				100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
Chernihiv	Urban	Line	143	107	161	215	67	79	97	107	120	155	195
		Rate		15.7	68.0	80.6	3.9	11.4	35.4	47.8	63.7	76.6	84.9
	Rural	Line	118	107	161	215	67	79	97	107	120	155	195
		Rate		28.3	60.5	73.0	14.5	23.4	34.8	45.0	49.5	63.0	86.1
	All	Line	261	107	161	215	67	79	97	107	120	155	195
		Rate		20.2	65.3	77.8	7.7	15.7	35.2	46.8	58.6	71.7	85.4
City of Kyiv	Urban	Line	326	107	161	215	67	79	97	107	120	155	195
		Rate		16.1	45.0	68.2	4.8	10.3	24.1	31.4	39.8	59.5	70.1
	Rural	Line	—	—	—	—	—	—	—	—	—	—	—
		Rate		—	—	—	—	—	—	—	—	—	—
	All	Line	326	107	161	215	67	79	97	107	120	155	195
		Rate		16.1	45.0	68.2	4.8	10.3	24.1	31.4	39.8	59.5	70.1

Poverty lines and poverty rates (spanning header over National and Percentile-based lines)

Source: 2018 HLCS.

Poverty rates are percentages.

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

Percentile-based lines are UAH per-person, per-day.

All poverty lines are UAH in prices in Ukraine as a whole on average during the 2018 HLCS 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.[15]

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:

$$\frac{5+4}{1+1} = \frac{9}{2} = 4.5 \text{ people.}$$

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 4.5 members. The estimated number of people in the households of in-coming participants is then 1,000 · 4.5 = 4,500.

The third and final step is to multiply the estimated poverty rate (here, 48.8 percent, or 0.488) by the estimated number of people in in-coming households (here, 4,500). This gives 4,500 · 0.488 ≈ 2,195 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,[16] 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.

---

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

[16] 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.

Two sampling approaches are possible for the follow-up round after baseline:

- *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,[17] 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.[18]

Continuing the earlier example, suppose that the first household at follow-up has four members (rather than five as at baseline) and is scored a second time on August 13, 2024, which is 1,157 days (about 3.17 years) after its first interview on June 13, 2021. Its score is now 36 (rather than 35), so its poverty likelihood for 100% of the national line remains unchanged at 58.2 percent (Figure 1).

Suppose that the second household now has six members (rather than four as at baseline) and is scored a second time on May 15, 2024, which is 1,050 days (about 2.88 years) after its first interview on June 30, 2021. Its score is now 40 (rather than 43), so its poverty likelihood has increased from 40.6 to 54.8 percent.

---

[17] Or when the follow-up sample is a random sample of the baseline sample.

[18] 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 +7.5 percentage points:

$$\frac{\left(\frac{5+4}{2}\right)\cdot(0.582-0.582)+\left(\frac{4+6}{2}\right)\cdot(0.548-0.406)}{\left(\frac{5+4}{2}\right)+\left(\frac{4+6}{2}\right)}\approx\frac{0.000+0.71}{9.5}\approx+0.075.$$

The estimated head-count poverty rate increased (improved) by 7.5 percentage points (not by 7.5 *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:

$$\frac{\left(\frac{5+4}{2}\right)\cdot 3.17+\left(\frac{4+6}{2}\right)\cdot 2.88}{\left(\frac{5+4}{2}\right)+\left(\frac{4+6}{2}\right)}\approx\frac{14.26+14.38}{9.5}\approx 3.02\text{ years.}$$

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: +7.5 ÷ 3.02 ≈ +2.5 percentage points per year.[19] The positive change means that poverty increased.[20]

In practice, the calculations should be done in an app or spreadsheet like Figure 11.

[19] *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.

[20] 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-2021	13-Aug-2024	3.17 = (C3-B3)/365	5	4	4.50 = (E3+F3)/2	14.26 = D3*G3	35	36	58.2	58.2	0 = G3*(L3-K3)/100
4	2W3120ZG8	30-Jun-2021	15-May-2024	2.88 = (C4-B4)/365	4	6	5.00 = (E4+F4)/2	14.38 = D4*G4	43	40	40.6	54.8	+0.71 = G4*(L4-K4)/100
5				Average:	4.5 = AVERAGE(E3:E4)	5.0 = AVERAGE(F3:F4)	Sum:	28.65 = SUM(H3:H4)					+0.71 = SUM(M3:M4)
6													
7						Estimated net change in head-count poverty rate (percentage points), follow-up versus baseline:							+7.5 = M5/(E5+F5)*100
8													
9								Household-size-weighted average years between interviews:					3.02 = H6/(E5+F5)
10													
11							Estimated annual net change in head-count poverty rate (percentage points):						+2.5 = 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:					4,000 = (E5*M13+F5*M14)/2
17													
18								Estimated annual net change in the number of poor people:					+99 = 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. Rather, the bottom line is 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 in 2021 of in-coming households in the calendar-year 2021 cohort was 1,000. By the end of the follow-up period of calendar-year 2024, 300 had dropped out, leaving 700. If drop-out took place at a constant pace and was unrelated to changes in poverty,[21] 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

---

[21] 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, 2016a). 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, 4.5 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), divided by the number of survey rounds (two). This is $\frac{4.5 \cdot 1{,}000 + 5.0 \cdot 700}{1+1} = 4{,}000$ people.

The second and last step is to multiply the estimated annual change in the poverty rate (here, about +2.5 percentage points, or +0.025) by the estimated average number of on-going participants (here, 4,000). This gives an estimate of the annual net change in the number of poor people by 100% of the national line of $+0.025 \cdot 4{,}000 \approx +99$ people.[22] This positive change is a an increase (worsening) in poverty; there are about 99 more 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 real 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 year, 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? After all, decision-making hinges on forecasts of the expected impacts of possible choices; 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

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

the program works (such as those in Figure 10). A program can also look for signs that participants value (or 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.[23]

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.[24] 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, 2024. It has three members, a score of 39, and a poverty likelihood by 100% of the national line of 58.2 percent (Figure 1).

The fourth household is interviewed on April 4, 2024. It has seven members, a score of 42, and a poverty likelihood of 54.8 percent.

At follow-up, the estimated head-count poverty rate is calculated in the same way as at baseline, that is, as the household-size-weighted average of the poverty likelihoods of the sampled households:

$$\frac{3 \cdot 0.582 + 7 \cdot 0.548}{3+7} \approx \frac{1.75 + 3.84}{10} \approx 0.558 = 55.8 \text{ percent.}$$

[23] Schreiner (2016a) and Schreiner (2014).

[24] 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 (55.8 percent) versus at baseline (50.4 percent),[25] divided by the difference (in years) between the household-size-weighted average of follow-up interview dates (March 25, 2024) versus the household-size-weighted average of baseline interview dates (June 10, 2021). These two average dates differ by about 1,009 days or about 2.76 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 100% percent of the national line, this is (55.8 – 50.4) ÷ 2.76 ≈ +2.0 percentage points per year.

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

[25] 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. Thus, the figure here is 50.4 percent, not 50.4 – 1.6 = 48.8 percent.

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-2021	5	07-Apr-2507 = C2*D2	35	58.2	2.91 = D2*G2/100
3	Baseline	2W3120ZG8	30-Jun-2021	4	30-Dec-2385 = C2*D2	43	40.6	1.62 = D3*G3/100
4	Follow-up	3XA76T21L	3-Mar-2024	3	09-Jul-2272 = C2*D2	39	58.2	1.75 = D4*G4/100
5	Follow-up	4Y8Y3EQS9	4-Apr-2024	7	01-Nov-2769 = C2*D2	42	54.8	3.84 = D5*G5/100
6		Sum baseline:		9 = SUM(D2:D3)				4.53 = SUM(H2:H3)
7		Sum follow-up:		10 = SUM(D4:D5)				5.58 = SUM(H4:H5)
8		Average baseline:		4.5 = AVERAGE(D2:D3)	20-Jun-2021 = SUM(E2:E3)/D6			
9		Average follow-up:		5.0 = AVERAGE(D4:D5)	25-Mar-2024 = SUM(E4:E5)/D7			
10								
11							Estimated baseline poverty rate (%):	50.4 = H6/D6*100
12							Estimated follow-up poverty rate (%):	55.8 = H7/D7*100
13								
14							Average years between follow-up and baseline interviews:	2.76 = (E9-E8)/365
15								
16							Estimated annual net change in head-count poverty rate (percentage points):	+2.0 = (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:	4,000 = (D8*H18+D9*H19)/2
22								
23							Estimated annual net change in the number of poor people:	+79 = 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 2021 cohort in 2021 is 1,000 in-coming households. By the end of the 2024 follow-up period, 300 households dropped out, leaving 700. If drop-out took place at a constant pace and was unrelated with 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, 4.5 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 $\frac{4.5 \cdot 1{,}000 + 5.0 \cdot 700}{1+1} = 4{,}000$ people.

The second and last step is to multiply the estimated annual net change in the head-count poverty rate (here, +2.0 percentage points, or +0.020) by the estimated number of on-going participants (here, 4,000). For 100% of the national line, this gives an annual net change in the number of poor people of +0.020 · 4,000 ≈ +79 people per year. This positive change is a (non-compounded) increase in poverty; the number of poor people in participating households increases (worsens) by 79 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:[26]

- 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 issues 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 Ukraine's SSSU in the 2018 HLCS, so it provides the most-accurate and most-consistent data (and thus the best estimates).

---

[26] 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[27] 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[28]

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.

[27] Schreiner, 2015.

[28] 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 or appropriate, 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 to be 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. This approach may also be less costly at follow-up, given that the sampled households have already been tracked down 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[29] 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 these 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.

---

[29] Schreiner, 2013.

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.[30]

Households that score at or below a given cut-off should be labeled as *targeted*,[31] not as *poor*.[32]

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*).

---

[30] *Targeting status* (having a score at or below a targeting cut-off) is not the same concept as *poverty status* (having income below a poverty line). Poverty status is a fact that is defined by whether income 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.

[31] 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 income 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*.

[32] After all, unless all targeted households have poverty likelihoods of 100 percent, it is likely that some of them are non-poor (their income is above a given poverty line). In the context of the scorecard, the terms *poor* and *non-poor* have specific, income-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

		Targeting segment	
		Targeted	Non-targeted
Observed poverty status	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.[33]

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

- 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.

---

[33] 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 100%	150%	200%	Percentile-based lines 10th	20th	40th	50th	60th	80th	90th
<=34	5.1	3.5	4.6	5.1	1.7	2.8	3.9	4.2	4.5	5.0	5.1
<=39	10.4	6.4	9.0	10.1	2.8	4.8	7.3	7.9	8.5	10.0	10.2
<=42	15.7	8.9	13.8	15.2	4.4	6.8	10.7	11.9	12.9	15.0	15.4
<=44	20.8	11.4	18.1	20.2	5.5	8.7	13.7	15.2	17.0	19.8	20.4
<=46	24.8	12.8	21.4	24.1	6.1	9.8	15.8	17.8	19.9	23.4	24.3
<=48	29.0	14.2	24.9	28.2	6.6	10.6	17.9	20.5	23.1	27.2	28.3
<=50	35.2	16.7	29.7	34.2	7.6	12.5	21.1	24.2	27.4	32.8	34.3
<=52	40.2	18.1	33.9	38.9	8.2	13.3	23.6	27.4	31.2	37.5	39.1
<=54	44.7	19.3	37.0	43.1	8.5	13.9	25.2	29.6	33.9	40.9	43.3
<=56	51.8	20.6	41.5	49.2	8.9	14.9	27.9	32.8	37.7	46.5	49.5
<=58	56.8	21.8	45.0	53.6	9.2	15.6	29.4	35.0	40.6	50.6	54.0
<=60	64.3	22.8	50.4	60.4	9.3	16.1	31.0	38.5	45.4	56.7	61.1
<=62	69.2	24.0	53.8	64.5	9.9	16.9	32.4	40.7	48.2	60.4	65.4
<=64	75.2	24.4	56.7	69.0	10.0	17.5	34.3	42.7	50.8	64.1	70.0
<=66	81.4	25.0	60.7	74.5	10.0	18.3	36.9	45.6	54.5	69.4	75.5
<=69	85.7	25.3	62.7	77.7	10.1	18.5	37.6	46.6	56.0	72.2	79.1
<=72	91.2	26.3	65.2	81.7	10.4	19.4	39.2	48.4	58.3	75.9	83.9
<=77	95.5	26.5	66.1	84.1	10.4	19.5	39.4	48.7	58.7	77.6	86.7
<=100	100.0	26.5	66.9	86.4	10.4	19.5	39.7	48.9	59.2	79.3	89.5

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			Percentile-based lines						
		100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
<=34	5.1	23.0	62.3	81.4	8.7	16.7	35.7	44.7	54.7	74.2	84.4
<=39	10.4	20.1	57.9	76.3	7.6	14.7	32.4	41.0	50.7	69.3	79.3
<=42	15.7	17.6	53.1	71.2	6.1	12.7	28.9	37.1	46.3	64.3	74.1
<=44	20.8	15.2	48.8	66.2	5.0	10.9	26.0	33.8	42.2	59.5	69.1
<=46	24.8	13.7	45.5	62.3	4.3	9.8	23.9	31.1	39.3	55.9	65.2
<=48	29.0	12.3	42.0	58.2	3.8	8.9	21.8	28.4	36.1	52.1	61.2
<=50	35.2	9.8	37.2	52.3	2.8	7.0	18.5	24.8	31.8	46.5	55.2
<=52	40.2	8.4	33.0	47.5	2.3	6.2	16.1	21.5	28.0	41.8	50.4
<=54	44.7	7.2	29.9	43.3	2.0	5.6	14.5	19.3	25.3	38.4	46.2
<=56	51.8	5.9	25.4	37.2	1.6	4.7	11.8	16.1	21.5	32.8	40.0
<=58	56.8	4.7	21.9	32.8	1.3	4.0	10.3	13.9	18.6	28.7	35.5
<=60	64.3	3.7	16.5	26.0	1.2	3.4	8.7	10.4	13.9	22.6	28.4
<=62	69.2	2.5	13.1	21.9	0.5	2.6	7.2	8.3	11.0	18.9	24.1
<=64	75.2	2.1	10.2	17.4	0.5	2.0	5.4	6.2	8.4	15.2	19.4
<=66	81.4	1.5	6.2	12.0	0.5	1.2	2.8	3.4	4.7	9.9	13.9
<=69	85.7	1.2	4.2	8.7	0.4	1.0	2.0	2.3	3.3	7.1	10.4
<=72	91.2	0.2	1.7	4.7	0.0	0.1	0.4	0.6	0.9	3.4	5.6
<=77	95.5	0.0	0.9	2.4	0.0	0.0	0.3	0.3	0.5	1.7	2.8
<=100	100.0	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			Percentile-based lines						
		100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
<=34	5.1	1.6	0.5	0.0	3.4	2.3	1.2	0.9	0.6	0.1	0.0
<=39	10.4	4.0	1.3	0.3	7.5	5.5	3.1	2.5	1.9	0.4	0.2
<=42	15.7	6.7	1.9	0.5	11.3	8.9	4.9	3.8	2.8	0.7	0.3
<=44	20.8	9.5	2.7	0.6	15.3	12.1	7.2	5.6	3.8	1.0	0.5
<=46	24.8	12.0	3.4	0.7	18.7	15.0	9.0	6.9	4.9	1.4	0.5
<=48	29.0	14.8	4.1	0.7	22.4	18.4	11.1	8.5	5.9	1.8	0.7
<=50	35.2	18.4	5.4	1.0	27.5	22.7	14.0	11.0	7.8	2.4	0.9
<=52	40.2	22.2	6.3	1.3	32.1	27.0	16.7	12.8	9.0	2.8	1.2
<=54	44.7	25.4	7.6	1.5	36.2	30.8	19.5	15.1	10.8	3.8	1.4
<=56	51.8	31.2	10.3	2.7	43.0	37.0	23.9	19.0	14.1	5.3	2.4
<=58	56.8	35.0	11.8	3.2	47.6	41.3	27.4	21.8	16.2	6.2	2.8
<=60	64.3	41.5	13.8	3.9	55.0	48.2	33.3	25.7	18.9	7.6	3.3
<=62	69.2	45.2	15.5	4.7	59.3	52.2	36.8	28.5	21.0	8.8	3.9
<=64	75.2	50.7	18.5	6.2	65.2	57.6	40.9	32.5	24.4	11.0	5.2
<=66	81.4	56.3	20.7	6.9	71.4	63.0	44.4	35.8	26.9	12.0	5.9
<=69	85.7	60.4	23.0	8.0	75.6	67.2	48.1	39.1	29.7	13.5	6.6
<=72	91.2	64.9	26.0	9.5	80.8	71.8	52.0	42.8	32.9	15.3	7.3
<=77	95.5	69.0	29.4	11.4	85.0	76.0	56.1	46.8	36.7	17.9	8.8
<=100	100.0	73.5	33.1	13.6	89.6	80.5	60.3	51.1	40.8	20.7	10.5

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			Percentile-based lines						
		100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
<=34	5.1	71.9	32.6	13.5	86.2	78.2	59.2	50.2	40.2	20.7	10.5
<=39	10.4	69.5	31.8	13.3	82.0	74.9	57.3	48.6	38.9	20.3	10.3
<=42	15.7	66.7	31.2	13.1	78.2	71.6	55.4	47.2	38.0	20.0	10.2
<=44	20.8	64.0	30.4	13.0	74.2	68.3	53.2	45.5	36.9	19.7	10.0
<=46	24.8	61.5	29.7	12.9	70.9	65.4	51.3	44.1	35.9	19.3	10.0
<=48	29.0	58.7	29.0	12.8	67.2	62.0	49.2	42.6	34.9	18.9	9.8
<=50	35.2	55.1	27.7	12.6	62.0	57.8	46.3	40.1	33.0	18.3	9.6
<=52	40.2	51.3	26.8	12.3	57.5	53.5	43.7	38.3	31.7	17.9	9.3
<=54	44.7	48.1	25.4	12.0	53.3	49.6	40.9	36.0	30.0	16.9	9.1
<=56	51.8	42.3	22.8	10.9	46.6	43.4	36.4	32.1	26.7	15.4	8.1
<=58	56.8	38.4	21.3	10.4	41.9	39.2	32.9	29.3	24.6	14.5	7.7
<=60	64.3	32.0	19.3	9.7	34.5	32.2	27.0	25.3	21.9	13.1	7.3
<=62	69.2	28.3	17.6	8.9	30.2	28.2	23.5	22.5	19.7	11.9	6.6
<=64	75.2	22.7	14.6	7.4	24.4	22.9	19.4	18.6	16.4	9.7	5.3
<=66	81.4	17.1	12.4	6.7	18.2	17.4	15.9	15.3	13.9	8.7	4.6
<=69	85.7	13.1	10.1	5.6	13.9	13.3	12.3	12.0	11.0	7.2	3.9
<=72	91.2	8.6	7.1	4.1	8.8	8.7	8.4	8.2	7.9	5.4	3.2
<=77	95.5	4.5	3.7	2.1	4.5	4.5	4.2	4.2	4.0	2.8	1.7
<=100	100.0	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			Percentile-based lines						
		100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
<=34	5.1	75.4	37.2	18.6	87.9	81.0	63.1	54.4	44.7	25.7	15.5
<=39	10.4	75.9	40.8	23.4	84.8	79.8	64.6	56.5	47.4	30.3	20.4
<=42	15.7	75.7	45.1	28.3	82.6	78.4	66.1	59.1	51.0	35.0	25.5
<=44	20.8	75.4	48.5	33.1	79.7	77.0	66.8	60.6	53.9	39.6	30.4
<=46	24.8	74.3	51.1	37.1	77.0	75.2	67.2	61.9	55.8	42.7	34.3
<=48	29.0	72.9	53.8	41.1	73.8	72.6	67.1	63.1	58.0	46.1	38.1
<=50	35.2	71.8	57.4	46.7	69.7	70.3	67.4	64.3	60.4	51.1	43.9
<=52	40.2	69.4	60.7	51.2	65.7	66.8	67.3	65.7	62.9	55.4	48.4
<=54	44.7	67.4	62.5	55.2	61.8	63.5	66.1	65.6	63.9	57.8	52.4
<=56	51.8	62.9	64.3	60.1	55.4	58.3	64.3	65.0	64.4	61.9	57.6
<=58	56.8	60.2	66.2	64.0	51.1	54.7	62.3	64.4	65.2	65.1	61.7
<=60	64.3	54.8	69.7	70.1	43.8	48.4	58.0	63.9	67.2	69.8	68.3
<=62	69.2	52.3	71.4	73.4	40.2	45.2	56.0	63.2	67.9	72.3	71.9
<=64	75.2	47.2	71.3	76.4	34.3	40.4	53.7	61.3	67.2	73.8	75.4
<=66	81.4	42.2	73.2	81.1	28.2	35.8	52.8	60.9	68.4	78.1	80.2
<=69	85.7	38.4	72.8	83.3	24.0	31.8	49.9	58.6	67.0	79.4	83.1
<=72	91.2	35.0	72.3	85.8	19.2	28.1	47.6	56.6	66.2	81.2	87.1
<=77	95.5	30.9	69.7	86.2	15.0	24.0	43.6	52.9	62.8	80.4	88.4
<=100	100.0	26.5	66.9	86.4	10.4	19.5	39.7	48.9	59.2	79.3	89.5

Scorecard applied to the validation sample.

For an example cut-off of 44 or less in the previous figures, 20.8 percent of all people are targeted, and outcomes for 100% of the national line in the validation sample are:

- Inclusion: 11.4 percent are below the line and correctly targeted
- Undercoverage: 15.2 percent are below the line and mistakenly not targeted
- Leakage: 9.5 percent are above the line and mistakenly targeted
- Exclusion: 64.0 percent are above the line and correctly not targeted

Increasing the cut-off to 46 or less increases the share of of all people targeted to 24.8 percent. The higher cut-off improves inclusion and undercoverage but worsens leakage and exclusion:

- Inclusion: 12.8 percent are below the line and correctly targeted
- Undercoverage: 13.7 percent are below the line and mistakenly not targeted
- Leakage: 12.0 percent are above the line and mistakenly targeted
- Exclusion: 61.5 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 100% of the national line in the validation sample, total net benefit under the hit rate for a cut-off of 44 or less is 75.4 percent. That is, about three in four Ukrainians 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			Percentile-based lines						
		100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
<=34	5.1	69.1	90.2	99.3	34.2	55.7	77.1	82.3	88.5	99.0	99.3
<=39	10.4	61.8	87.4	97.5	27.2	46.7	70.5	76.3	82.1	96.5	97.9
<=42	15.7	57.0	88.1	96.8	27.8	43.5	68.5	75.5	82.5	95.6	97.8
<=44	20.8	54.6	86.9	97.0	26.3	41.7	65.6	73.0	81.6	95.3	97.7
<=46	24.8	51.6	86.2	97.4	24.7	39.4	63.8	72.0	80.2	94.4	97.9
<=48	29.0	49.0	85.7	97.4	22.8	36.5	61.7	70.8	79.8	93.7	97.6
<=50	35.2	47.6	84.6	97.2	21.7	35.5	60.1	68.8	77.9	93.3	97.4
<=52	40.2	44.9	84.3	96.7	20.3	33.0	58.6	68.2	77.5	93.1	97.1
<=54	44.7	43.2	82.9	96.6	18.9	31.1	56.4	66.3	75.9	91.5	96.8
<=56	51.8	39.8	80.1	94.9	17.1	28.6	53.8	63.4	72.8	89.7	95.4
<=58	56.8	38.3	79.2	94.4	16.1	27.4	51.7	61.7	71.5	89.1	95.0
<=60	64.3	35.5	78.5	94.0	14.4	25.1	48.2	60.0	70.6	88.2	94.9
<=62	69.2	34.7	77.7	93.2	14.3	24.5	46.9	58.8	69.6	87.3	94.3
<=64	75.2	32.5	75.4	91.8	13.3	23.3	45.6	56.8	67.6	85.3	93.1
<=66	81.4	30.7	74.6	91.5	12.3	22.5	45.4	56.0	67.0	85.3	92.8
<=69	85.7	29.5	73.1	90.7	11.7	21.6	43.9	54.4	65.3	84.2	92.3
<=72	91.2	28.9	71.5	89.6	11.4	21.3	43.0	53.0	63.9	83.2	92.0
<=77	95.5	27.7	69.2	88.0	10.9	20.5	41.3	51.0	61.5	81.2	90.8
<=100	100.0	26.5	66.9	86.4	10.4	19.5	39.7	48.9	59.2	79.3	89.5

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			Percentile-based lines						
		100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
<=34	5.1	2.2:1	9.2:1	132.6:1	0.5:1	1.3:1	3.4:1	4.7:1	7.7:1	100.5:1	132.6:1
<=39	10.4	1.6:1	6.9:1	38.8:1	0.4:1	0.9:1	2.4:1	3.2:1	4.6:1	27.2:1	47.0:1
<=42	15.7	1.3:1	7.4:1	30.5:1	0.4:1	0.8:1	2.2:1	3.1:1	4.7:1	21.9:1	44.9:1
<=44	20.8	1.2:1	6.7:1	32.1:1	0.4:1	0.7:1	1.9:1	2.7:1	4.4:1	20.3:1	42.9:1
<=46	24.8	1.1:1	6.3:1	36.9:1	0.3:1	0.7:1	1.8:1	2.6:1	4.1:1	16.8:1	47.5:1
<=48	29.0	1.0:1	6.0:1	37.9:1	0.3:1	0.6:1	1.6:1	2.4:1	3.9:1	15.0:1	39.9:1
<=50	35.2	0.9:1	5.5:1	34.2:1	0.3:1	0.6:1	1.5:1	2.2:1	3.5:1	13.8:1	37.5:1
<=52	40.2	0.8:1	5.4:1	29.7:1	0.3:1	0.5:1	1.4:1	2.1:1	3.5:1	13.5:1	33.1:1
<=54	44.7	0.8:1	4.8:1	28.2:1	0.2:1	0.5:1	1.3:1	2.0:1	3.1:1	10.7:1	30.3:1
<=56	51.8	0.7:1	4.0:1	18.5:1	0.2:1	0.4:1	1.2:1	1.7:1	2.7:1	8.7:1	20.6:1
<=58	56.8	0.6:1	3.8:1	16.8:1	0.2:1	0.4:1	1.1:1	1.6:1	2.5:1	8.1:1	19.0:1
<=60	64.3	0.5:1	3.6:1	15.6:1	0.2:1	0.3:1	0.9:1	1.5:1	2.4:1	7.5:1	18.7:1
<=62	69.2	0.5:1	3.5:1	13.7:1	0.2:1	0.3:1	0.9:1	1.4:1	2.3:1	6.9:1	16.6:1
<=64	75.2	0.5:1	3.1:1	11.2:1	0.2:1	0.3:1	0.8:1	1.3:1	2.1:1	5.8:1	13.5:1
<=66	81.4	0.4:1	2.9:1	10.8:1	0.1:1	0.3:1	0.8:1	1.3:1	2.0:1	5.8:1	12.8:1
<=69	85.7	0.4:1	2.7:1	9.7:1	0.1:1	0.3:1	0.8:1	1.2:1	1.9:1	5.3:1	12.0:1
<=72	91.2	0.4:1	2.5:1	8.6:1	0.1:1	0.3:1	0.8:1	1.1:1	1.8:1	4.9:1	11.5:1
<=77	95.5	0.4:1	2.2:1	7.4:1	0.1:1	0.3:1	0.7:1	1.0:1	1.6:1	4.3:1	9.9:1
<=100	100.0	0.4:1	2.0:1	6.4:1	0.1:1	0.2:1	0.7:1	1.0:1	1.5:1	3.8:1	8.5:1

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			Percentile-based lines						
		100%	150%	200%	10th	20th	40th	50th	60th	80th	90th
<=34	5.1	13.3	6.9	5.9	16.7	14.5	9.9	8.6	7.6	6.4	5.7
<=39	10.4	24.1	13.5	11.7	27.0	24.8	18.4	16.2	14.3	12.6	11.3
<=42	15.7	33.7	20.7	17.6	41.7	35.0	27.1	24.2	21.8	18.9	17.2
<=44	20.8	42.8	27.0	23.4	52.3	44.5	34.4	31.0	28.7	25.0	22.7
<=46	24.8	48.3	32.0	27.9	58.7	50.1	39.9	36.4	33.6	29.5	27.2
<=48	29.0	53.5	37.2	32.7	63.4	54.2	45.1	41.9	39.0	34.3	31.6
<=50	35.2	63.1	44.4	39.5	73.1	64.0	53.3	49.4	46.2	41.4	38.3
<=52	40.2	68.1	50.7	45.0	78.3	68.1	59.4	56.0	52.7	47.3	43.7
<=54	44.7	72.7	55.3	49.9	81.0	71.1	63.5	60.5	57.2	51.5	48.4
<=56	51.8	77.7	62.0	56.9	84.7	76.1	70.3	67.1	63.7	58.6	55.3
<=58	56.8	82.1	67.2	62.0	87.8	79.8	74.1	71.6	68.6	63.8	60.4
<=60	64.3	86.0	75.4	69.9	88.6	82.5	78.1	78.7	76.6	71.5	68.3
<=62	69.2	90.7	80.4	74.7	95.0	86.7	81.8	83.1	81.4	76.2	73.0
<=64	75.2	92.1	84.7	79.8	95.5	89.7	86.4	87.2	85.8	80.9	78.3
<=66	81.4	94.3	90.8	86.2	95.6	93.8	93.1	93.1	92.1	87.5	84.4
<=69	85.7	95.4	93.7	89.9	96.4	94.6	94.9	95.3	94.5	91.0	88.4
<=72	91.2	99.4	97.5	94.6	99.7	99.3	98.9	98.8	98.5	95.7	93.8
<=77	95.5	99.9	98.7	97.3	100.0	100.0	99.3	99.4	99.2	97.8	96.9
<=100	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, 70 percent—among targeted people. For 100% of the national line, targeting Ukrainians who score 34 or less would target 5.1 percent of people in Ukraine and give a head-count poverty rate among those targeted of 69.1 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 100% of the national line and a score cut-off of 34 or less, 2.2 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 Ukrainians. Figure 21 shows that a score cut-off of 48 or less would target 29.0 percent of all Ukrainians, a group in which 53.5 percent are poor by 100% of the national line.

Interview Guide

The excerpts quoted here are from:

SSSU. (2020) *Statistical Yearbook of Ukraine: 2019*, [the *Yearbook*], link.

_____. (2011) "Методологічні Положення З Організації Державного Статистичного Спостереження *Обстеження Умов Життя Домогосподарств*", [the *Manual*], 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 Ukraine 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".

Likewise, do not directly ask the third scorecard question ("Is the head of the household male?"). Instead, mark the response based on what you the enumerator already know about the household head from when you compiled the "Back-page Worksheet".

Finally, do not directly ask the fourth scorecard question ("How many members of the household are wage/salary workers?"). Instead, mark the response based on the number of household members that you recorded as wage/salary employees on the "Back-page Worksheet".

Ask all of the six 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 participant of record 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 participant of record 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 blank spaces in the header. 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. This 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. Do not read the response options.

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:

5. How many rooms does the household's residence have?	A. One	0	
	B. Two	5	5
	C. Three or more	7	

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 Ukraine’s SSSU in the 2018 HLCS. 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 HLCS by Ukraine’s SSSU. For example, interviews should done in-person by a trained enumerator at the residence of the participating household because that is what the SSSU did in the 2018 HCLS.

G2. Translation

You the enumerator should do the interview in a language which both you and the respondent speak and understand well.

The scorecard itself, the "Back-page Worksheet", and this "Guide" are available in English and Ukrainian. There are not yet official, professional translations to other languages spoken in Ukraine. 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 to arrange to collaborate on one.

G3. General interview guidance from the *Manual*

G3.1 Who should be the respondent?

According to page 24 of the *Manual*, "The respondent may be any adult member of the household who is able to provide the requested information. The preferred respondent in this group is the one who best knows the household's budget (expenses and income). This is usually a woman (the head of the household if the head is a woman, or the wife of the male head of the household, if the head is a man). Ideally, you the enumerator should do the interview in the presence both of this woman and of the male head of the household."

G3.2 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 participant of record with your program (although the head may be the participant of record).

According to the first page of the 2018 HLCS questionnaire "Control Card for Household Composition", the *head of the household* is "the member of the household who is in charge of maintaining and running the household. This is determined by the household members themselves. In case of disagreement or uncertainty among the household members, a common rule of thumb is that the head is the person who earns the most income."

G4. Guidelines for each question in the scorecard

G4.1 In which region does the household live?

A. Kharkiv, or Ivano-Frankivsk
B. Cherkasy, Volyn, or Sumy
C. City of Kyiv, Kyiv (excluding the city of Kyiv), Kherson, Rivne, or Chernivtsi
D. Zakarpattya, or Khmelnytskiy
E. Odesa, Zaporizhzhya, Vinnytsya, Poltava, Mykolayiv, or Luhansk
F. Donetsk, Zhytomyr, or Kirovohrad
G. Lviv, Dnipropetrovsk, Chernihiv, or Ternopil

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. Four or more
B. Three
C. Two
D. 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. 72 of the *Yearbook* and pages 19 and 20 of the *Manual*, “a *household* is one person alone or a group of people—regardless of blood or marital relationships—who live together in a residence, who keep house together, and who partly or fully share funds. Members of the household include people who live permanently with the household, even if they are temporarily absent on the day of the interview (as long as their total expected absence is 12 months or less).

“Household members do not need to be be related by blood or marriage. For example, a group of unrelated students may be members of a single household if they rent a residence together and share expenses for food.

“To qualify as a member of the interviewed household, a person must meet both of the following conditions:

- He/she considers the residence of the interviewed household to be his/her only or main place of residence
- He/she shares (partly or fully) the expenses for food and shelter with all other household members”

According to pages 19 and 20 of the *Manual*, “Do not count as a member of the interviewed household any person—regardless of his/her relationship with members of the interviewed household—who who usually lives in this household but who has been absent (or who is expected to be absent) from the household for a total duration of 12 months or more. Examples of such persons—assuming that the total actual or expected duration of their absence is 12 months or more—include those who are:

- Hospitalized
- Incarcerated
- Performing military service
- Living elsewhere as a student
- On a business trip
- Working elsewhere

“For example, suppose that the interviewed household includes parents who have a son who is a student and who usually lives elsewhere at the place of his studies. Even if the son is in the residence of his parents on the day of the survey, he is not counted as a member of the interviewed household.

“In contrast, a person who is temporarily absent from the interviewed household on the day of the interview does count as a member of the interviewed household as long as he/she normally resides in the household and as long as the actual or expected total duration of the absence is less than 12 months. Example of temporarily absent members of the household include those who are:

- Doing temporary or seasonal work
- On internships
- On vacations, holidays, or business trips
- Visiting relatives or friends
- Absent from their usual residence for a total expected duration of less than 12 months because they are:
  - At hospitals, rehabilitation facilities, maternity hospitals, and the like
  - Students at boarding schools (except those who live there permanently)
  - Conscripts called up for military training
  - Detainees awaiting trial
  - Under arrest (if the total expected duration of their absence from the usual residence is less than 12 months)

“If a person has been (for example) serving in the military, working elsewhere on a contract, or in prison, then he/she should not be counted as a member of the interviewed household if he/she is not expected to return before the total duration of his/her absence exceeds 12 months. However, if he/she is expected to return soon enough that the total duration of his/her absence will be less than 12 months, then he/she should be counted as temporarily absent and thus included as a member of the interviewed household.

“If a student studies elsewhere yet regularly returns to the interviewed household (for example, on weekends or on Sundays), then he/she is considered to be temporarily absent and thus is counted as a member of the interviewed household.

“In contrast, if a student studies elsewhere and visits the interviewed household only for holidays or vacation periods, then he/she is not temporarily absent and is not counted as a member of the interviewed household. However, if a student is expected to complete his/her studies and then to return to live in the interviewed household within the next 12 months, then he/she is considered to be temporarily absent and is counted as a member of the interviewed household.”

According to page 36 of the *Manual*, “Do not count as members of the interviewed household anyone who is staying temporarily with the interviewed household in its residence, even if these visitors (such relatives, friends, or people requiring medical care) stay for a long time and even if they share some of the costs of food and shelter.”

G4.3 Is the head of the household male?

E. No
F. Yes

Do not directly ask this question of the respondent. Instead, mark the response based on the sex of the head of the household that you the enumerator recorded as part of the "Back-page Worksheet".

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

According to the first page of the 2018 HLCS questionnaire "Control Card for Household Composition", the *head of the household* is "the member of the household who is in charge of maintaining and running the household. This is determined by the household members themselves. In case of disagreement or uncertainty among the household members, a common rule of thumb is that the head is the person who earns the most income."

G4.4 How many members of the household are wage/salary employees?

A. None
B. One
C. Two or more

Do not directly ask this question of the respondent. Instead, mark the response based on the count of wage/salary employees that you have already recorded as part of the “Back-page Worksheet”.

The *Manual* does not define what a “wage/salary employee” is.

G4.5 How many rooms does the household's residence have?

A. One
B. Two
C. Three or more

According to page 39 of the *Manual*, "A room is a space intended for occupation that is separated from other spaces by partitions. For the purposes of this question, kitchens, halls, corridors, bathrooms, showers, pantries, built-in closets, and other ancillary rooms in the apartment or house are not counted as rooms.

"Insulated attics, mezzanines, terraces, and verandas that are finished and suitable for year-round use count as rooms for the purposes of this question.

"Count all the rooms in the residence of the interviewed household, even if the interviewed household rents out some of the rooms to another household."

G4.6 Does the household's residence have hot water?

A. No
B. Yes

According to pages 40 to 42 of the *Manual*, "This question seeks to determine the presence of hot water, regardless of who paid for its installation.

"Take care in cases in which there is hot water that works intermittently, weakly, or not at all. In some regions, there is hot water, but it is only available intermittently, or there is no hot water despite the fact that the physical infrastructure is in place to provide this service to apartments and houses. In these cases, you the enumerator should count the interviewed household's residence as having hot water as long as the infrastructure is in place, even if the service itself is intermittent, weak, or non-functional.

"A household is considered to have hot water if there is a system of pipes that provides hot water for baths, showers, laundry, washing dishes, and other household or business uses. To have hot water requires that there be a system (centralized or local) that carries hot water to all places in the residence in which hot water is used (kitchen, bathroom, toilet, and other places). For example, the residence of the interviewed household is not considered to have hot water if its hot water comes from its own hot-water heater (not a system) that is installed in single place in the residence (such as a gas or electric water heater in the bathroom) which does not also send hot water through pipes to the kitchen, toilet, and possibly other places. If the residence has a centralized hot water system, and if the residence of the interviewed household does not actually receive and use the hot water from this system (for example, because the interviewed household uses its own hot water heater), the residence nevertheless is considered to have hot water (as long as it has its own system within the residence, as outlined above)."

G4.7 Does the household's residence have sewer service?

A. No
B. Yes

According to pages 40 to 42 of the *Manual*, "This question seeks to determine the presence of sewer service, regardless of who paid for its installation.

"Take care in cases in which there is sewer service that works intermittently, weakly, or not at all. In some regions, there is sewer service, but it is only available intermittently, or there is no sewer service despite the fact that the physical infrastructure is in place to provide this service to apartments and houses. In these cases, you the enumerator should count the interviewed household's residence as having sewer service as long as the infrastructure is in place, even if the service itself is intermittent, weak, or non-functional.

"*Sewer service* is a system of drainage that carries waste water and human waste away from the residence into a septic tank or into a public network of pipes or channels designed for that purpose."

G4.8 Does the household have a clothes-washing machine?

A. No
B. Yes

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

G4.9 Does the household have a microwave oven?

C. No
D. Yes

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

G4.10 In the past three months, did the household keep any farm animals of its own, such as livestock, poultry, bees, and so on?

A. No
B. Yes

According to page 45 of the *Manual*, "Only count livestock, poultry, bees, and so on that are owned by the interviewed household.

"Do not count livestock, poultry, bees and so on that the interviewed household does not own but rather takes care of on contract on behalf of an agricultural firm or on behalf of another household."

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

The State Statistics Service of Ukraine (SSSU) fielded the 2018 Household Living Conditions Survey (HLCS) with 8,039 households from October to November of 2018. The 2018 HLCS is Ukraine's most-recent available national household income survey that includes data on ownership of consumer durables.

Questions and points for the scorecard are selected (*constructed*) based on data from a random three-fifths of the 8,039 households in the 2018 HLCS. 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 HLCS 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 income (UAH 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 income from the 2018 HCLS. SSSU (2020, p. 73) describes the measurement of income in the HCLS.

Because pro-poor programs in Ukraine may want to use different or various poverty lines, the scorecard supports 10 lines:

- 100% of the national line
- 150% of the national line
- 200% of the national line
- First-decile (10th-percentile) line
- First-quintile (20th-percentile) line
- Second-quintile (40th-percentile) line
- Median (50th-percentile) line
- Third-quintile (60th-percentile) line
- Fourth-quintile (80th-percentile) line
- Tenth-decile (90th-percentile) line

A2.1 National poverty lines

For scorecard purposes, Ukraine's national poverty line (called here "100% of the national line") is taken as the SSSU's non-official "actual subsistence minimum", defined in terms of income (SSSU, 2017). This line is about UAH107 per adult equivalent per day in average prices for Ukraine as a whole during the 2018 HLCS fieldwork, giving a head-count poverty rate of 26.7 percent (Figure 10).[34]

[34] This rate matches World Bank (2020), suggesting that this paper uses the same data and calculations.

Like most official poverty lines around the world, this non-official line for Ukraine is based on the cost-of-basic-needs method that reflects the cost of a minimum standard for food and non-food.[35] It is adjusted over time for changes in prices, so its poverty-rate estimates from the scorecard and based on this line can be compared across years. This non-official line (used here) is to be preferred over Ukraine's official poverty line whose estimates cannot be compared over time because its constant-price value changes year-to-year based on governmental policy decisions.[36] The official line also gives an unrealistically low head-count poverty rate (1.3 percent).

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

A2.2 International 2011 PPP poverty lines

The World Bank tracks world-wide poverty with four 2011 PPP poverty lines:[37]

- $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

The scorecard does not support any of these 2011 PPP lines. The head-count poverty rate in Ukraine for the three lower lines is 4 percent or less; for the highest line, it is 89 percent. Pro-poor programs in Ukraine can safely assume that more or less none of their participants are below $5.50/day, and that more or less all are below $21.70/day.

[35] Ravallion, 1998.
[36] World Bank, 2020.
[37] Jolliffe and Prydz, 2016; Ferreira *et al.*, 2016.

A2.3 Percentile-based poverty lines

The scorecard does support percentile-based poverty lines.[38] This facilitates a number of types of analyses. For example, the second-quintile (40th-percentile) line might be used to help track Ukraine'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 income with health outcomes (or anything else related with the distribution of income). 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.[39]

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

- Relative wealth (via scores)
- Absolute income (via poverty likelihoods and absolute poverty lines)
- Relative income (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 (income 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*.

---

[38] Percentiles are defined in terms of all people in Ukraine. For example, the all-Ukraine head-count poverty rate for the first-quintile (20th-percentile) poverty line is 20.0 percent (Figure 10: (Ukraine overall, Vinnytsya, and Volyn): Poverty lines and head-count poverty rates by urban/rural/all in 2018
).

[39] Rutstein and Johnson, 2004.

Annex 3 Scorecard construction

For Ukraine, about 50 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))
- Employment (such as the number of household members who are wage/salary employees)
- Housing (such as the number of rooms or the presence of sewer service)
- Ownership of consumer durables (such as microwave ovens or clothes-washing machines)
- Agriculture (such as ownership of land plots or livestock)
- Recent food consumption (such as non-alcoholic beverages other than water or milk)
- 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 non-alcoholic beverages other than water or milk is probably more responsive to changes in poverty than is the age of the head of the household).

The scorecard itself is built using 100% 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.[40]

[40] Ravallion, 2009.

One of the one-question draft scorecards is then selected based on:[41]

- Improvement in accuracy
- Acceptability to users in terms of:
  - Simplicity
  - Cost of collection
  - Concordance with:
    - Experience
    - Theory
    - Common sense
- Sensitivity to changes in income
- 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 income
- 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[42] 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.

---

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

[42] 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 Ukraine. Customizing poverty-assessment tools by urban/rural does not improve targeting accuracy much.[43] Segment-specific tools may improve the accuracy of estimates of poverty rates,[44] but:

- They run a greater risk of overfitting[45]
- 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 Ukraine, the region of residence)[46]

[43] 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.
[44] Diamond *et al.*, 2016; Tarozzi and Deaton, 2009.
[45] Haslett, 2012.
[46] Schreiner, 2016b.

Annex 4 Estimates of poverty likelihoods

This annex tells how scores are converted into estimates of 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 do not stand for 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 100% of the national line, scores of 43–44 correspond with a poverty likelihood of 40.6 percent, and scores of 49–50 correspond with a poverty likelihood of 35.1 percent (Figure 1).

The poverty likelihood associated with a score varies by poverty line. For example, scores of 43–44 are associated with a likelihood of 40.6 percent for 100% of the national line but with a likelihood of 25.1 percent for the first-quintile (20th-percentile) 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 income below a given poverty line.

For the example of 100% of the national line and a score of 43–44 (Figure 22 below), there are 3,653 (normalized) households in the construction sample. Of these, 1,484 (normalized) are below the poverty line. The estimated poverty likelihood associated with a score of 43–44 is then 40.6 percent, because 1,484 ÷ 3,653 ≈ 0.406 = 40.6 percent.

The same method is used to calibrate all scores with poverty likelihoods for all 10 poverty lines.[47]

---

[47] 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 (100% of national line)

Score	Households in range and < poverty line		All households in range		Poverty likelihood (%)
0–34	2.594	÷	3.613	=	71,8
35–39	3.452	÷	5.927	=	58,2
40–42	3.147	÷	5.738	=	54,8
43–44	1.484	÷	3.653	=	40,6
45–46	1.790	÷	4.407	=	40,6
47–48	2.407	÷	5.925	=	40,6
49–50	2.309	÷	6.576	=	35,1
51–52	1.596	÷	5.996	=	26,6
53–54	1.571	÷	5.900	=	26,6
55–56	1.516	÷	5.696	=	26,6
57–58	1.020	÷	5.207	=	19,6
59–60	854	÷	5.948	=	14,4
61–62	533	÷	4.569	=	11,7
63–64	467	÷	5.593	=	8,3
65–66	374	÷	4.519	=	8,3
67–69	283	÷	4.742	=	6,0
70–72	321	÷	5.494	=	5,8
73–77	345	÷	5.901	=	5,8
78–100	142	÷	4.595	=	3,1

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 income.[48] 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 $2.718281828^{\beta X} \times (1 + 2.718281828^{\beta X})^{-1}$, 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 Ukraine's 2018 HLCS who are below a poverty line. A calibrated look-up table also allows analysts to convert scores to likelihoods without any math at all. This calibration approach can also improve accuracy, especially with large samples.

48 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 Ukraine's scorecard for snapshot estimates of head-count poverty rates in a single time period by 100% of the national poverty line is +1.6 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 50.4 percent and the error is +1.6 percentage points, then an improved estimate is 50.4 – (+1.6) = 48.8 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 10 poverty lines.

Errors are derived *out-of-sample*. This means that the scorecard (made from the construction sample from the 2018 HLCS, 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 income-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 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.[49] 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.[50]

It is possible to compare estimated and observed poverty rates because the validation sample from the 2018 HLCS records actual (not estimated) income-based poverty status. The observed poverty likelihood in the 2018 HLCS 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.

---

[49] Estimation errors due to being out-of-time can be measured with post-2018 data (say, from a future HLCS). Of course, future HLCS 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).

[50] 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 Ukraine overall. The population's true value is taken as the value in the 2018 HLCS (even though the HLCS 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.[51] 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.[52]

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 Ukraine's 10 poverty lines.

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 scorecard 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.[53]

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.

---

[51] 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.
[52] Schreiner (2021) discusses the derivation of errors.
[53] 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 getting at least six "heads" (60 percent of the 10 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.

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 100% of the national line is 48.8 percent and that the sample size is $n = 1,024$. Given 90-percent confidence,[54] the margin of error is ±2.6 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 from 48.8 – 2.6 = 46.2 percent to 48.8 + 2.6 = 51.4 percent, with the most-likely true value being the center of the range (the 48.8-percent estimate).

Said another way, "With 90-percent confidence, the estimate has a margin of error from 46.2 to 51.4 percent." This means that the true population value has a:

- 5-percent chance of being less than 46.2 percent
- 90-percent chance of being between 46.2 and 51.4 percent
- 5-percent chance of being greater than 51.4 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 Ukraine probably is indeed pro-poor if the scorecard estimate of the poverty rate for in-coming participants by 100% of the national poverty line with 80-percent confidence is 35.0 percent with a margin of error of ±5.0 percentage points, that is, from 30.0 to 40.0 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-Ukraine rate for this line of 26.7 percent from Figure 10.

[54] 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 20.0 to 50.0 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 Ukrainian (26.7 percent) 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 Ukraine 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 ±3.3 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 ±3.3 percentage points or less of the error-adjusted estimate.

A5.2.4 How to calculate margins of error

The spreadsheet-based PovIt!™ reporting app calculates margins of error for all scorecard estimates discussed here. Analysts may also use the formulas below.[55]

[55] Schreiner (2021) 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, $\sqrt{\dfrac{\hat{p} \cdot (1-\hat{p})}{n}} \cdot \varphi$,

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

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

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[56] 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, Ukraine's 2018 HLCS gives a direct-measure head-count poverty rate for 100% of the national line of $\hat{p}$ = 26.7 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.[57] Ukraine in 2018 had a population of households (not people) of N = 14,920,040, and the HLCS sample size was n = 8,039. 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:

$$\pm z \cdot \alpha \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \pm 1.64 \cdot 1.00 \cdot \sqrt{\frac{0.267 \cdot (1-0.267)}{8{,}039}} \cdot \sqrt{\frac{14{,}920{,}040 - 8{,}039}{14{,}920{,}040 - 1}}.$$

This implies a 90-percent chance that Ukraine's true head-count poverty rate for 100% of the national line in 2018 is in the range of 26.7 – 0.8 = 25.9 percent to 26.7 + 0.8 = 27.5 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.

[56] 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.

[57] For scorecard estimates, α for a given poverty line is found in Figure 2.

To illustrate, the baseline example in Section 3 has an estimated snapshot poverty rate of 48.8 percent. With 70-percent confidence, the margin of error is about ±45.6 percentage points,[58] or from 48.8 – 45.6 = 3.2 percent to 48.8 + 45.6 = 94.4 percent. The margin of error is huge because the sample size of $n = 2$ interviewed households is very small.[59]

The estimated number of people in participating households in the example in Section 3 is 7,000,[60] so the lower limit of the 70-percent margin of error for the estimated number of poor people is 4,500 · 0.032 = 144. The upper limit is 4,500 · 0.944 = 4,248. This example estimate—based as it is on a sample of two households—is useless, because it is consistent with almost none or almost all of people in participating households being poor.

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

---

[58] The example in Section 3 has $N = 1{,}000$, $n = 2$, and $\alpha = 1.24$ (Figure 2). For 70-percent confidence, $z = 1.04$. The margin of error $\pm c$ for the head-count poverty-rate estimate is then $\pm 0.456 \approx \pm 1.04 \cdot 1.24 \cdot \sqrt{\frac{0.488 \cdot (1 - 0.488)}{2}} \cdot \sqrt{\frac{1{,}000 - 2}{1{,}000 - 1}}$.

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

[60] 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 rows 3 and 4 of column M in Figure 11, here zero (because all changes are positive), divided by the sum across all sampled households of each household's average household size across baseline and follow-up of 4.5 + 5.0 = 9.5 (from row 5, columns E and F). Thus, $\hat{p}_{up} = 0.00 \div 9.5 = 0.000$.

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 rows 3 and 4 of column M in Figure 11), which is (+0.00) + (+0.71) = +0.71. Dividing this by the sum across all sampled households of each household's average household size across baseline and follow-up (4.5 + 5.0 = 9.5) gives $\hat{p}_{down} = 0.71 \div 9.5 \approx 0.075$.[61]

The household-size-weighted average of the number of years between interviews y is 3.02 (from row8, column M in Figure 11).

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,[62] the margin of error $\pm c$ is about ±0.073 ≈

$$\pm \frac{1.04 \cdot 1.14}{3.02} \cdot \sqrt{\frac{0.000 \cdot (1-0.000) + 0.075 \cdot (1-0.075) + 2 \cdot 0.000 \cdot 0.075}{2}} \cdot \sqrt{\frac{1{,}000-2}{1{,}000-1}}.$$

The example's estimated net annual poverty-rate change is +2.5 percentage points (Figure 11), so the 70-percent margin of error is from +2.5 – 7.3 = –4.8 to +2.5 + 7.3 = +9.8 percentage points. The estimate from this tiny sample of $n = 2$ is uninformative; the true net change could easily be strongly negative, close to zero, or strongly positive.

This example shows why margins of error are useful. Without them, program managers might believe that there was evidence that poverty rates increased by 2.5

---

[61] $\hat{p}_{down} - \hat{p}_{up}$ is the estimated net poverty-rate change. In this particular example, $\hat{p}_{down} \approx 0.075$ and $\hat{p}_{up} = 0.000$, so $0.075 - 0.000 \approx +0.075$, which indeed is the estimated 7.5 percentage-point increase in the poverty rate in Figure 11.

[62] Schreiner, 2021.

percentage points per year even though the data in this sample is also consistent with widely different rates and directions of change.

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 3.3.1 for one sample scored twice, the estimated annual net change in the poverty rate is +2.5 percentage points. As just shown, the tiny sample size of $n = 2$ means that the 70-percent margin of error runs from –4.8 to +9.8 percentage points.

The estimated average number of on-going participating people is 4,000.[63] Thus, the lower limit of the 70-percent margin of error for the estimated annual net change in the number of poor people is $4{,}000 \cdot (-0.048) \approx -192$ (a net decrease in poor people), and the upper limit is $4{,}000 \cdot (+0.098) \approx +392$ (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.

[63] 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 example for two independent samples in Section 3.3.4:

- $z = 1.04$, assuming a desired confidence level is 70 percent
- $\alpha = 1.10$, the adjustment factor (regardless of poverty line) for this estimator[64]
- $y = 2.76$, the years between the average interview at baseline and follow-up
- $\hat{p} = 0.504$, 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

The margin of error $\pm c$ is $\pm 0.207 \approx \pm \frac{1.04 \cdot 1.10}{2.76} \cdot \sqrt{\frac{2 \cdot 0.504 \cdot (1 - 0.504)}{2}} \cdot \sqrt{\frac{850 - 2}{850 - 1}}$.

The example's estimated net annual poverty-rate change is +2.0 percentage points (Figure 12). Thus, the 70-percent margin of error is from +2.0 – 20.7 = –18.7 percentage points to +2.0 + 20.7 = +22.7 percentage points. The tiny sample is again consistent with a true value in the population that is strongly negative, close to zero, or strongly 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 3.3.4 for two independent samples estimates the annual net change in the poverty rate as +2.0 percentage points. As just shown, the 70-percent margin of error runs from –18.7 to +22.7 percentage points.

The estimated average number of on-going participating people is 4,000.[65] 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 4,000 · (–0.187) ≈ –748 (a net decrease in poor people), and the upper limit is 4,000 · (+0.227) ≈ +908 (a net increase in poor people).

---

[64] Schreiner, 2021.

[65] 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 (after planning for reporting and analysis) 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[66]
- 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.

[66] 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 $n = N \cdot \left(\dfrac{z^2 \cdot \alpha^2 \cdot \tilde{p} \cdot (1 - \tilde{p})}{z^2 \cdot \alpha^2 \cdot \tilde{p} \cdot (1 - \tilde{p}) + c^2 \cdot (N - 1)} \right)$,

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.[67]

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 100% of the national line, so α = 1.24 (Figure 2)
- The pre-estimation expected poverty rate is the all-Ukraine rate for 100% of the national line in 2018, so $\tilde{p}$ = 26.7 percent = 0.267 (Figure 10)
- The desired margin of error ±c = ±3.0 percentage points = ±0.030

Given these hypothetical values,

$$n = 10{,}000 \cdot \left(\frac{1.28^2 \cdot 1.24^2 \cdot 0.267 \cdot (1 - 0.267)}{1.28^2 \cdot 1.24^2 \cdot 0.267 \cdot (1 - 0.267) + 0.03^2 \cdot (10{,}000 - 1)} \right) \approx 520.$$

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

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:[68]

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

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 100% of the national line
- The desired confidence level for the margin of error is 80 percent, so $z = 1.28$
- α = 1.14 (regardless of the scorecard or poverty line[69]
- 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-Ukraine rate for 100% of the national line: $p_{\text{pre-baseline}}$ = 26.7 percent = 0.267 (Figure 10)
- The population of participating households is $N = 10{,}000$

Assuming N is large relative to n so that $\sqrt{\frac{N-n}{N-1}} \approx 1$, then the baseline sample size n is $2 \cdot \left(\frac{1.28 \cdot 1.14}{0.03}\right)^2 \cdot [-0.01 + 0.016 \cdot 3 + 0.56 \cdot 0.267 \cdot (1 - 0.267)] \cdot 1 \approx 699.$

The follow-up sample size is also 699.

68 Schreiner, 2021.
69 Schreiner, 2021.

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 $n = 2 \cdot N \cdot \left(\frac{z^2 \cdot \alpha^2 \cdot \tilde{p} \cdot (1-\tilde{p})}{z^2 \cdot \alpha^2 \cdot \tilde{p} \cdot (1-\tilde{p}) + c^2 \cdot (N-1)} \right)$.[70]

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.[71]

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:

$$2 \cdot 10{,}000 \cdot \left(\frac{1.28^2 \cdot 1.10^2 \cdot 0.267 \cdot (1-0.267)}{1.28^2 \cdot 1.10^2 \cdot 0.267 \cdot (1-0.267) + 0.03^2 \cdot (10{,}000-1)} \right) \approx 827.$$

The sample size at follow-up is also n = 827.

---

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

[71] Schreiner, 2021.

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.

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. (2021) "Scorocs Simple Poverty Scorecard Tool: Uganda", link.

_____. (2016a) "Popular Science", link.

_____. (2016b) "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.

_____. (2013) "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.

State Statistics Service of Ukraine. (2020) *Statistical Yearbook of Ukraine*: 2019, link.

_____. (2017) "Principles in Organizing the Monitoring of Poverty and Effectiveness of Social Programmes in Ukraine", link.

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) “Poverty and Equity Brief: Ukraine”, 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.