Simple Poverty Scorecard® Poverty-Assessment Tool Palestine (West Bank and Gaza Strip)

Shiyuan Chen, Mark Schreiner, and Gary Woller

September 9, 2008

This document is at `SimplePovertyScorecard.com`.

Abstract

The Simple Poverty Scorecard-brand poverty-assessment tool uses 10 low-cost indicators from the 2005 Palestine (West Bank and Gaza Strip) Expenditure and Consumption Survey to estimate the likelihood that a household has consumption below a given poverty line. Field workers can collect responses in about ten minutes. The scorecard's accuracy is reported for a range of poverty lines. The scorecard is a practical way for pro-poor programs in Palestine (West Bank and Gaza Strip) to measure poverty rates, to track changes in poverty rates over time, and to segment clients for differentiated treatment.

Acknowledgements

Simple Poverty Scorecard® Poverty-Assessment Tool

Interview ID:		**Name**	**Identifier**
Interview date:		Participant:	
Country:	PSE	Field agent:	
Scorecard:	001	Service point:	
Sampling wgt.:		Number of household members:	

Indicator	Response	Points	Score
1. How many household members are 18-years-old or younger?	A. Six or more	0	
	B. Five	5	
	C. Three or four	9	
	D. Two	16	
	E. One	30	
	F. None	33	
2. What is the household's main source of income?	A. Subsistence farming, household business, wages and salaries from private sector, or social affairs	0	
	B. Wages and salaries from public sector or Israeli sector, or cash remittances received from within Palestine	7	
	C. Cash remittances received from abroad, international organization, or other	13	
3. Does the household have a private car?	A. No	0	
	B. Yes	14	
4. Does the household have a phone line?	A. No	0	
	B. Yes	7	
5. Does the household have a satellite dish?	A. No	0	
	B. Yes	4	
6. What is the main source of energy for heating?	A. Coal/firewood, electricity, none, or no data	0	
	B. Gas, kerosene, or other	7	
7. Does the household have a solar boiler?	A. No	0	
	B. Yes	5	
8. What type of connection to sewage networks does the household have?	A. Cesspit, or no connection to sewage system	0	
	B. Public sewage system	5	
9. Does the household have a refrigerator?	A. No	0	
	B. Yes	8	
10. Does the household live in a refugee camp?	A. No	0	
	B. Yes	4	

Score:

Simple Poverty Scorecard® Poverty-Assessment Tool Palestine (West Bank and Gaza Strip)

1. Introduction

Pro-poor programs in Palestine (West Bank and Gaza Strip) can use the Simple Poverty Scorecard poverty-assessment tool to estimate the likelihood that a household has consumption below a given poverty line, to estimate a population's poverty rate at a point in time, to track changes in a population's poverty rate over time, and to segment participants for differentiated treatment.

The direct approach to poverty measurement via direct surveys is difficult and costly, asking households about a lengthy list of consumption items (such as "In the past month, how much did you spend on bread and cereals?", or "In the past month, how much did you spend on clothing and footwear?").

In contrast, the indirect approach via the scorecard is quick and inexpensive. It uses 10 verifiable indicators (such as "What is the household's main source of income?" or "Does the household have a private car?") to get a score that is highly correlated with poverty status as measured by the exhaustive survey.

The scorecard here differs from "proxy means tests" (Coady, Grosh, and Hoddinott, 2002) in that it is tailored to the capabilities and purposes not of national governments but rather of local, pro-poor organizations. The feasible poverty-measurement options for these organizations are typically subjective and relative (such

as participatory wealth ranking by skilled field workers) or blunt (such as rules based on land-ownership or housing quality). Results from these approaches are not comparable across organizations nor across countries, they may be costly, and their accuracy is unknown.

If an organization wants to know what share of its participants are below a poverty line (say, $1/day for the Millennium Development Goals, or the poorest half below the national poverty line as required of USAID microenterprise grantees), or if it wants to measure movement across a poverty line (for example, to report to the Microcredit Summit Campaign), then it needs an consumption-based, objective tool with known accuracy. While consumption surveys are costly even for governments, many small, local organizations can implement an inexpensive scorecard that can serve for monitoring, management, and targeting.

The statistical approach here aims to be understood by non-specialists. After all, if managers are to adopt the scorecard on their own and apply it to inform their decisions, they must first trust that it works. Transparency and simplicity build trust. Getting "buy-in" matters; proxy means tests and regressions on the "determinants of poverty" have been around for three decades, but they are rarely used to inform decisions, not because they do not work, but because they are presented (when they are presented at all) as tables of regression coefficients incomprehensible to lay people (with cryptic indicator names such as "`HHSIZE_2`", negative values, many decimal places, and standard errors). Thanks to the predictive-modeling phenomenon known as the "flat

max", simple, transparent scorecards can be almost as accurate as complex, opaque ones.

The technical approach here is also innovative in how it associates scores with poverty likelihoods, in the extent of its accuracy tests, and in how it derives sample-size formulas. Although these techniques are simple and/or standard, they have rarely or never been applied to proxy means tests.

The scorecard is based on the 2005 PECS conducted by the Palestinian Central Bureau of Statistics (PCBS). Indicators are selected to be:

- Inexpensive to collect, easy to answer quickly, and simple to verify
- Strongly correlated with poverty
- Liable to change over time as poverty status changes

All points in the scorecard are non-negative integers, and total scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line). Non-specialists can collect data and tally scores on paper in the field in about five minutes.

The scorecard can be used to estimate three basic quantities. First, it can estimate a household's "poverty likelihood", that is, the probability that the household has per-capita consumption below a given poverty line.

Second, the scorecard can estimate the poverty rate of a group of households at a point in time. This is simply the average poverty likelihood among the households in the group.

Third, the scorecard can estimate changes in the poverty rate for a group of households between two points in time. This estimate is simply the change in the average poverty likelihood of the households in the group over time.

The scorecard can also be used for targeting. To help managers choose a targeting cut-off, this paper reports the share of Palestine's households who are below a given poverty line and who are also at or below a given score cut-off.

This paper presents a single scorecard whose indicators and points are derived from Palestine household consumption data and the national poverty line. Scores from this scorecard are calibrated to poverty likelihoods for six poverty lines.

The scorecard is constructed using a sub-sample of the data from the 2005 PECS. Its accuracy is validated on a different sub-sample from the 2005 PECS. While all three scoring estimators are unbiased when applied to the validation sample (that is, they match the true value on average in repeated samples from the same population from which the scorecard was built), they are—like all predictive models—biased to some extent when applied to a different population.

Thus, while the indirect scoring approach is less costly than the direct survey approach, it is also biased. (The survey approach is unbiased by assumption.) There is bias because scoring must assume that the future relationship between indicators and

poverty will be the same as in the data used to build the scorecard.[1] Of course, this assumption—ubiquitous and inevitable in predictive modeling—holds only partly.

When applied to the validation sample, the absolute difference between scorecard estimates of groups' poverty rates and the true rates is 0.1 percentage points for the national line, and 1.0 percentage points on average across all six lines. This difference is due to sampling variation and not bias; its average would be zero if the whole PECS were to be repeatedly redrawn and divided into sub-samples before repeating the entire scorecard-building process.

For sample sizes of $n = 16,384$, the 90-percent confidence intervals for these estimates are ±0.5 or less. For $n = 1,024$, the 90-percent intervals are ±2.0 percentage points or less.

Section 2 below describes data and poverty lines. Sections 3 and 4 describe scorecard construction and offer practical guidelines for use. Sections 5 and 6 detail the estimation of households' poverty likelihoods and of groups' poverty rates at a point in time. Section 7 discusses estimating changes in poverty rates. Section 8 covers targeting. The final section is a summary.

---

[1] Bias may also result from changes in the quality of data collection, from imperfect adjustment of poverty lines across time or geographic regions, or from sampling variation across consumption surveys.

2. Data and poverty lines

This section discusses the data used to construct and test the scorecard. It also presents the poverty lines to which scores are calibrated.

2.1   Data

The scorecard is based on data from the 2005 PECS. The data is randomly divided into three sub-samples (Figure 2):

- *Construction* for selecting indicators and points
- *Calibration* for associating scores with poverty likelihoods
- *Validation* for testing accuracy on data not used in construction or calibration

2005 was a "normal" period between crises in Palestine. The scorecard here is thus most appropriately applied to other similar non-crisis periods.

2.2 Household consumption and expenditure

The 2005 PECS (Central Bureau of Statistics, 2006) calculates what it calls

"household consumption", defined as the sum of:

Bread and cereals	Salt, spices and other food	Transport and communication
Meat and poultry		
Fish and sea products	Take-away food and meals in restaurants	Education
Dairy products and eggs		Recreation
Oils and fats	Own-produced food in-kind	Personal care
Fruits and nuts		Tobacco
Vegetables, legumes and tubers	Clothing and footwear	Alcohol beverage
	Housing	Other non-food consumption
Sugar and confectionery	Furniture and utensils	
Non-alcoholic beverages	Household operations	Other than food
	Medical care	Imputed rent

The 2005 PECS also calculates a value that it calls "household expenditure"

defined as "household consumption" from above, plus:

Remittances in cash
Taxes in cash
Other cash non-consumption expenditure
Social security

minus

Own produced food in kind
Other than food
Imputed rent

Poverty status can be defined in terms of "consumption" or "expenditure". With

the 2005 PECS and the national poverty line, the household-level poverty rate based on

"consumption" is 18.4 percent, and it is 26.8 percent based on "expenditure" (Figure 16).

The corresponding person-level poverty rates are 23.6 and 32.2.

Which definition is best? According to Deaton and Zaidi (2002), a measure of household aggregate consumption should include:

- Food consumption
- Non-food consumption
- Imputed housing rent
- Imputed use-value of durable goods
- Education expenditures
- Expenditures on water and electricity

This measure should exclude:

- Taxes (unless they are "fees for services" that directly benefit the household)
- Debt repayment
- Interest payments on debt
- Large and infrequent expenses such as marriages and dowries
- Health expenditures
- Gifts, transfers, and remittances set out

Neither "consumption" nor "expenditure" as calculated in the 2005 PECS data follows Deaton and Zaidi(2002). Both measures exclude the imputed use-value of consumer durables, both exclude expenditure on water and electricity, and both include health expenditures. In addition, the "expenditure" measure excludes imputed rent and the value of own-produce food in-kind, while including taxes and remittances sent out.

On net, the "consumption" measure in the 2005 PECS is closer to the Deaton/Zaidi ideal than the "expenditure" measure. From this point on, when this paper refers to *consumption*, it refers to what the 2005 PECS calls "consumption".

2.3 Poverty rates and poverty lines

2.3.1 Rates

There are two poverty rates, person-level and household-level. The person-level rate ("head-count index") is the share of people in a given group who live in households whose per-capita consumption (that is, total household consumption divided by the number of household members) is below a given poverty line.

The household-level poverty rate is the share of households in a given group whose per-capita consumption is below a given poverty line.

Whereas governments report person-level poverty rates, local pro-poor development organizations typically report household poverty rates. This is because development organizations want to know the poverty rate of their participants, not the poverty rate of all people who live in households with their participants. Thus, the household-level rate will typically be the benchmark when comparing the poverty rate of an organization's participants with the overall rate in a political entity.

Given household-level poverty likelihoods, the person-level poverty rate for all people in the group of households is simply the average of the household-level poverty likelihoods, weighted by the number of people in each household. Larger households are more likely to be poor, so the person-level rate usually exceeds the household-level rate.

2.3.2 Lines

The national poverty line developed by the Palestine National Poverty Commission in 1998 for West Bank and Gaza Strip (World Bank, 2004) corresponds to

9.86 New Israeli Shekels (NIS) per person per day in 2003 for the benchmark family of two adults and four children. This national poverty line is higher than the national 'subsistence' poverty line of NIS 6.74 per person per day, representing the cost of a minimum-calorie diet plus an allowance for basic non-food items.[2] The study here updates the national and 'subsistence' poverty line to 2005 with the Consumer Price Index (CPI) from the PCBS[3], to NIS 10.55 in 2005 and NIS 7.21 in 2005 (Figure 3).

The scorecard here is constructed using the national poverty line, adjusted for cost-of-living by region (West Bank or Gaza Strip) using regional CPIs as deflators. The national line produces a household-level poverty rate of 18.4 percent and a person-level poverty rate of 23.6 percent (Figure 3).

Because local pro-poor organizations may want to use different or various poverty lines, this paper calibrates scores from its single scorecard to poverty likelihoods for six lines (figures in parentheses are per-capita daily poverty lines in units of NIS, with household-level and person-level poverty rates from Figure 3)

- National line   (10.55,   18.4 percent, 23.6 percent)
- 'Subsistence' line   (7.21,   6.1 percent, 8.4 percent   )
- USAID "extreme" line   (8.06,   9.1 percent, 12.0 percent)
- $1/day   (5.55,   2.5 percent, 3.5 percent   )
- $2/day   (11.11,   20.7 percent, 26.5 percent)
- $3/day   (16.66,   43.9 percent, 52.8 percent)

---

[2] There is also a "food" line, but it is so low that very few Palestinians are below it.
[3] http://www.pcbs.gov.ps/Portals/_pcbs/cpi/2ed1a37d-d1d3-4f63-ae9e-18c7b9cc057d.htm, accessed September 10, 2008.

The USAID "extreme" line (U.S. Congress, 2002) is the median consumption of households below the national line.

The $1/day line is derived using:[4]

- 1993 purchase-power parity exchange rate: NIS2.54 per $1
- 1993 CPI for Israel: 72.7
- 1996 CPI for Palestine and Israel: 100
- 2005 CPI for Palestine: 146.8

The $1/day line for 2005 then (Sillers, 2006) is:

$$2.54 \times (146.8 \div 72.7) \times 1.08 = \text{NIS5.54}$$

The $1/day line is adjusted for differences in cost-of-living by region using:

- L, the all-Palestine $1/day line in NIS
- p_i, population proportions by region ($i = 1$ for West Bank, $i = 2$ for Gaza Strip)
- π_i, CPI by region.

L_i is then the $1/day line adjusted for regional cost-of-living:

$$L_i = \frac{L \cdot \pi_i}{\displaystyle\sum_{j=1}^{2} p_j \cdot \pi_j}.$$

The lines for $2/day and $3/day are multiples of the $1/day lines L_i.

---

3. Scorecard construction

About 200 potential indicators are initially prepared in the areas of:

- Family composition (such as household size and female headship)
- Employment (such as primary activity)
- Housing (such as type of dwelling and main flooring material)
- Ownership of durable goods (such as televisions and refrigerators)

Each indicator is first screened with the entropy-based "uncertainty coefficient" (Goodman and Kruskal, 1979) that measures how well it predicts poverty on its own. Figure 4 lists the best indicators, ranked by uncertainty coefficient. Responses for each indicator are ordered starting with those most strongly associated with poverty.

The scorecard also aims to measure *changes* in poverty through time. This means that, when selecting indicators and holding other considerations constant, preference is given to more sensitive indicators. For example, ownership of a television or a cooking stove is probably more likely to change in response to changes in poverty than is the education of the male head/spouse.

The scorecard itself is built using Logit regression on the construction sub-sample (Figure 2). Indicator selection uses both judgment and statistics (forward stepwise based on "c"). The first step is to build one scorecard for each candidate indicator, using Logit to derive points. Each scorecard's accuracy is taken as "c", a measure of ability to rank by poverty status (SAS Institute Inc., 2004).

One of these one-indicator scorecards is then selected based on several factors (Schreiner *et al.*, 2004; Zeller, 2004), including improvement in accuracy, likelihood of

acceptance by users (determined by simplicity, cost of collection, and "face validity" in terms of experience, theory, and common sense), sensitivity to changes in poverty status, variety among indicators, and verifiability.

A series of two-indicator scorecards are then built, each based on the one-indicator scorecard selected from the first step, with a second candidate indicator added. The best two-indicator scorecard is then selected, again based on "c" and judgment. These steps are repeated until the scorecard has 10 indicators.

The final step is to transform the Logit coefficients into non-negative integers such that total scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line).

This algorithm is the Logit analogue to the familiar R^2-based stepwise with least-squares regression. It differs from naïve stepwise in that the criteria for selecting indicators include not only statistical accuracy but also judgment and non-statistical factors. The use of non-statistical criteria can improve robustness through time and, more important, helps ensure that indicators are simple and make sense to users.

The single scorecard here applies to all of Palestine (West Bank and Gaza Strip). Evidence from India and Mexico (Schreiner, 2006 and 2005a), Sri Lanka (Narayan and Yoshida, 2005), and Jamaica (Grosh and Baker, 1995) suggests that segmenting poverty-assessment tools by urban/rural does not improve accuracy much.

4. Practical guidelines for scorecard use

The main challenge of scorecard design is not to squeeze out the last drops of accuracy but rather to improve the chances that scoring is actually used (Schreiner, 2005b). When scoring projects fail, the reason is not usually technical inaccuracy but rather the failure of an organization to decide to do what is needed to integrate scoring in its processes and to learn to use it properly (Schreiner, 2002). After all, most reasonable scorecards predict tolerably well, thanks to the empirical phenomenon known as the "flat max" (Hand, 2006; Baesens *et al.*, 2003; Lovie and Lovie, 1986; Kolesar and Showers, 1985; Stillwell, Hutton, and Edwards, 1983; Dawes, 1979; Wainer, 1976; Myers and Forgy, 1963). The bottleneck is less technical and more human, not statistics but organizational change management. Accuracy is easier to achieve than adoption.

The scorecard here is designed to encourage understanding and trust so that users will adopt it and use it properly. Of course, accuracy matters, but it is balanced against simplicity, ease-of-use, and "face validity". Programs are more likely to collect data, compute scores, and pay attention to the results if, in their view, scoring does not make a lot of "extra" work and if the whole process generally seems to make sense.

To this end, the scorecard here fits on one page. The construction process, indicators, and points are simple and transparent. "Extra" work is minimized; non-specialists can compute scores by hand in the field because the scorecard has:

- Only 10 indicators
- Only categorical indicators
- Simple weights (non-negative integers, no arithmetic beyond addition)

The scorecard in is ready to be photocopied. A field worker using the paper scorecard would:

- Record participant identifiers
- Read each question from the scorecard
- Circle the response and its points
- Write the points in the far-right column
- Add up the points to get the total score
- Implement targeting policy (if any)
- Deliver the paper scorecard to a central office for filing or data entry

Of course, field workers must be trained. Quality results depend on quality inputs. If organizations or field workers gather their own data and have an incentive to exaggerate poverty rates (for example, if they are rewarded for higher poverty rates), then it is wise to do on-going quality control via data review and random audits (Matul and Kline, 2003).[5] IRIS Center (2007a) and Toohig (2007) are useful nuts-and-bolts guides for budgeting, training field workers and supervisors, logistics, sampling, interviewing, piloting, recording data, and quality control.

In terms of sampling design, an organization must make choices about:

- Who will do the scoring
- How scores will be recorded
- What participants will be scored
- How many participants will be scored
- How frequently participants will be scored
- Whether scoring will be applied at more than one point in time
- Whether the same participants will be scored at more than one point in time

[5] If an organization does not want field workers to know the points associated with indicators, then they can use the version of Figure 1 without points and apply them later in a spreadsheet or database at the central office.

The non-specialists who apply the scorecard with participants in the field can be:

- Employees of the organization
- Third-party contractors

Responses, scores, and poverty likelihoods can be recorded:

- On paper in the field and then filed at an office
- On paper in the field and then keyed into a database or spreadsheet at an office
- In portable electronic devices in the field and downloaded to a database

The subjects to be scored can be:

- All participants (or all new participants)
- A representative sample of all participants (or of all new participants)
- All participants (or all new participants) in a representative sample of branches
- A representative sample of all participants (or of all new participants) in a representative sample of branches

If not determined by other factors, the number of participants to be scored can be derived from sample-size formulas (presented later) for a desired level of confidence and a desired confidence interval.

Frequency of application can be:

- At in-take only (precluding measuring change in poverty rates)
- As a once-off project for current participants (precluding measuring change)
- Once a year (or at some other fixed interval, allowing measuring change)
- Each time a field worker visits a participant at home (allowing measuring change)

When the scorecard is applied more than once in order to measure change in poverty rates, it can be applied:

- With a different set of participants
- With the same set of participants

An example set of choices were made by BRAC and ASA, two microlenders in Bangladesh (each with 7 million participants) who are applying the Simple Poverty Scorecard tool for Bangaldesh (Schreiner, 2013). Their design is that loan officers in a random sample of branches score all participants each time they visit a homestead as part of their standard due diligence prior to loan disbursement (about once a year). Responses are recorded on paper in the field before being sent to a central office to be entered into a database. ASA's and BRAC's sampling plans cover 50,000–100,000 participants each.

5. Estimates of household poverty likelihoods

The sum of scorecard points for a household is called the *score*. For Palestine, scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line). While higher scores indicate less likelihood of being below a poverty line, the scores themselves have only relative units. For example, doubling the score does not double the likelihood of being above a poverty line.

To get absolute units, scores must be converted to *poverty likelihoods*, that is, probabilities of being below a poverty line. This is done via simple look-up tables. For the example of the national line, scores of 0–4 have a poverty likelihood of 100 percent, and scores of 45–49 have a poverty likelihood of 6.4 percent (Figure 5).

The poverty likelihood associated with a score varies by poverty line. For example, scores of 45–49 are associated with a poverty likelihood of 6.4 percent for the national line but 34.6 percent for the \$3/day line.[6]

5.1 Calibrating scores with poverty likelihoods

A given score is non-parametrically associated ("calibrated") with a poverty likelihood by defining the poverty likelihood as the share of households in the calibration sub-sample who have the score and who are below a given poverty line.

---

[6] Starting with Figure 5, most figures have six versions, one for each poverty line. To keep them straight, they are grouped by poverty line. Single tables that pertain to all poverty lines are placed with the tables for the national line.

For the example of the national line (Figure 6), there are 5,726 households in the calibration sub-sample with a score of 20–24, of whom 3,036 are below the poverty line. The estimated poverty likelihood associated with a score of 20–24 is then 53.0 percent, because 3,036 ÷ 5,726 = 53.0 percent.

To illustrate with the national line and a score of 45–49, there are 8,888 households in the calibration sample, of whom 568 are below the line (Figure 6). Thus, the poverty likelihood for this score is 568 ÷ 8,888 = 6.4 percent.

The same method is used to calibrate scores with estimated poverty likelihoods for the other poverty lines.

Figure 7 shows, for all scores, the likelihood that consumption falls in a range demarcated by two adjacent poverty lines. For example, the daily consumption of someone with a score of 35–39 falls in the following ranges with probability:

- 0.0 percent   below $1/day
- 2.0 percent   between $1/day and the "subsistence" line
- 2.0 percent   between the "subsistence" line and the USAID "extreme" line
- 12.6 percent   between the USAID "extreme" line and the national line
- 2.8 percent   between the national line and $2/day
- 35.8 percent   between $2/day and $3/day
- 44.8 percent   above $3/day

Even though the scorecard is constructed partly based on judgment, the calibration process produces poverty likelihoods that are objective, that is, derived from data on consumption-based poverty lines. The poverty likelihoods would be objective even if indicators and/or points were selected without any data at all. In fact, objective scorecards of proven accuracy are often based only on judgment (Fuller, 2006; Caire,

2004; Schreiner *et al.*, 2004). Of course, the scorecard here is constructed with both data and judgment. The fact that this paper acknowledges that some choices in scorecard construction—as in any statistical analysis—are informed by judgment in no way impugns the objectivity of the poverty likelihoods, as this depends on using data in score calibration, not on using data (and nothing else) in scorecard construction.

Although the points in Palestine's scorecard are transformed coefficients from a Logit regression, scores are not converted to poverty likelihoods via the Logit formula of $2.718281828^{\text{score}}$ x $(1+ 2.718281828^{\text{score}})^{-1}$. This is because the Logit formula is esoteric and difficult to compute by hand. Non-specialists find it more intuitive to define the poverty likelihood as the share of households with a given score in the calibration sample who are below a poverty line. In the field, converting scores to poverty likelihoods requires no arithmetic at all, just a look-up table. This non-parametric calibration can also improve accuracy, especially with large calibration samples.

5.2   Accuracy of estimates of poverty likelihoods

As long as the relationship between indicators and poverty does not change, this calibration process produces unbiased estimates of poverty likelihoods. *Unbiased* means that in repeated samples from the same population, the average estimate matches the

true poverty likelihood. The scorecard also produces unbiased estimates of poverty rates at a point in time and of changes in poverty rates between two points in time.[7]

Of course, the relationship between indicators and poverty changes with time, so the scorecard applied after 2000 (as all are in practice) will generally be biased.

How accurate are estimates of poverty likelihoods? To measure, the scorecard is applied to 1,000 bootstrap samples of size $n = 16,384$ from the validation sub-sample (Figure 2). Bootstrapping entails:[8]

- Score each household in the validation sample
- Draw a new sample *with replacement* from the validation sample
- For each score, compute the true poverty likelihood in the bootstrap sample, that is, the share of households with the score and consumption below a poverty line
- For each score, record the difference between the estimated poverty likelihood (Figure 5) and the true poverty likelihood in the bootstrap sample
- Repeat the previous three steps 1,000 times
- For each score, report the average difference between estimated and true poverty likelihoods across the 1,000 bootstrap samples
- For each score, report the two-sided interval containing the central 900, 950, or 990 differences between estimated and true poverty likelihoods

For each score range, Figure 8 shows the average difference between estimated and true poverty likelihoods as well as confidence intervals around the differences.

For the national line, the average poverty likelihood across bootstrap samples for scores of 10–14 in the validation sample is too high by 0.7 percentage points (Figure 8). For scores of 15–19, the estimate is too low by 10.5 percentage points.[9]

---

[7] This follows because these estimates of groups' poverty rates are linear functions of the unbiased estimates of households' poverty likelihoods.
[8] Efron and Tibshirani, 1993.

For the validation sample, the 90-percent confidence interval for the differences for scores of 10–14 is ±3.7 percentage points (Figure 8).[10] This means that in 900 of 1,000 bootstraps, the difference between the estimate and the true value is between –3.0 and 4.4 percentage points (because 0.7 − 3.7 = –3.0, and 0.7 + 3.7 = 4.4). In 950 of 1,000 bootstraps (95 percent), the difference is 0.7 ±4.4 percentage points, and in 990 of 1,000 bootstraps (99 percent), the difference is 0.7 ±5.9 percentage points.

For almost all score ranges, Figure 8 shows differences—sometimes large ones—between estimated poverty likelihoods and true values. This is because the validation sub-sample is a single sample that—thanks to sampling variation—differs in distribution from the construction/calibration sub-samples and from Palestine's population. For targeting, however, what matters is less the bias in all score ranges and more the bias in score ranges just above and below the targeting cut-off. This fact mitigates the effects of bias and sampling variation on targeting (Friedman, 1997). Section 9 below looks at targeting accuracy in detail.

Of course, if estimates of groups' poverty rates are to be usefully accurate, then errors for individual households must largely cancel out. As discussed later, this is generally what happens.

---

[9] There are differences, in spite of the estimator's unbiasedness, because the estimates come from a single sample. Their average difference would be zero if samples were repeatedly drawn from the population and split into sub-samples before repeating the entire scorecard-building process.

[10] Confidence intervals are a standard, widely understood measure of precision.

Figure 9 (summarizing Figure 10 by poverty line) shows that absolute differences in the validation sample, when averaged across score ranges for a poverty line, are typically 1.4 percentage points or less, The differences are due to sampling variation.

By construction, the scorecard here is unbiased. It may still, however, be *overfit* when applied after 2005. That is, it may fit the 2005 PECS data so closely that it captures not only some timeless patterns but also some random patterns that, due to sampling variation, show up only in the 2005 PECS. Or the scorecard may be overfit in that it becomes biased as the relationship between indicators and poverty changes.

Overfitting can be mitigated by simplifying the scorecard and by not relying only on data but rather also considering experience, judgment, and theory. Of course, the scorecard here does this. Bootstrapping can also mitigate overfitting by reducing (but not eliminating) dependence on a single sampling instance. Combining scorecards can also help, at the cost of greater complexity.

Most errors in individual households' likelihoods, however, cancel out in the estimates of groups' poverty rates (see later sections). Furthermore, much of the differences may come from non-scorecard sources such as changes in the relationship between indicators and poverty, sampling variation, changes in poverty lines, inconsistencies in data quality across time, and inconsistencies in cost-of-living adjustments. These factors can be addressed only by improving data quantity and quality (which is beyond the scope of the scorecard) or by reducing overfitting (which likely has limited returns, given the scorecard's parsimony).

6. Estimates of a group's poverty rate at a point in time

A group's estimated poverty rate at a point in time is the average of the estimated poverty likelihoods of the individual households in the group.

To illustrate, suppose a program samples three households on Jan. 1, 2008 and that they have scores of 20, 30, and 40, corresponding to poverty likelihoods of 53.0, 32.0, and 10.7 percent (national line, Figure 5). The group's estimated poverty rate is the households' average poverty likelihood of $(53.0 + 32.0 + 10.7) \div 3 = 31.9$ percent.[11]

6.1   Accuracy of estimated poverty rates at a point in time

How accurate is this estimate? For a range of sample sizes, Figure 12 reports average differences between estimated and true poverty rates as well as precision (confidence intervals for the differences) for the scorecard applied to 1,000 bootstrap samples from the validation sample. For the national line, the scorecard is generally too low by about 0.1 percentage points; it estimates a poverty rate of 18.3 percent for the validation sample, but the true value is 18.4 percent (Figure 2). For all poverty lines, differences for the validation sample are 2.4 percentage points or less, with an average of about 1.0 percentage points (Figure 11).[12]

---

[11] The group's poverty rate is *not* the poverty likelihood associated with the average score. Here, the average score is $(20 + 30 + 40) \div 3 = 30$, and the poverty likelihood associated with the average score is 32.0 percent. This is not the 31.9 percent found as the average of the three poverty likelihoods associated with each of the three scores.
[12] Figure 11 summarizes Figure 12 across all poverty lines.

As before, these differences are due to sampling variation in the validation sample and in the random division of the 2005 PECS into three sub-samples.

In terms of precision, the 90-percent confidence interval for a group's estimated poverty rate at a point in time and $n = 16{,}384$ is 0.4 percentage points or less (Figure 11). This means that in 900 of 1,000 bootstraps of this size, the difference between the estimate and the true value is within 0.4 percentage points of the average difference. In the specific case of the national line and the validation sample, 90 percent of all samples of $n = 16{,}384$ produce estimates that differ from the true value in the range of –0.1 – 0.4 = –0.5 to –0.1 + 0.4 = 0.3 percentage points. (–0.1 is the average difference, and ±0.4 is its 90-percent confidence interval.)

6.2 Sample-size formula for estimates of poverty rates at a point in time

How many households should an organization sample if it wants to estimate their poverty rate at a point in time for a desired confidence interval and confidence level? This practical question was first addressed in Schreiner (2008a).[13]

---

[13] IRIS Center (2007a and 2007b) says that $n = 300$ is sufficient for USAID reporting. If a poverty-assessment tool is as precise as direct measurement, if the expected (before measurement) poverty rate is 50 percent, and if the confidence level is 90 percent, then $n = 300$ implies a confidence interval of ± 2.2 percentage points. In fact, USAID has not specified confidence levels or intervals. Furthermore, the expected poverty rate may not be 50 percent, and the tool could be more or less precise than direct measurement.

With direct measurement, the poverty rate can be estimated as the number of households observed to be below the poverty line, divided by the number of all observed households. The formula for sample size n in this case is (Cochran, 1977):

$$n = \left(\frac{z}{c}\right)^2 \cdot \hat{p} \cdot (1 - \hat{p}), \tag{1}$$

where

$$z \quad \text{is} \quad \begin{cases} 1.64 \text{ for confidence levels of 90 percent} \\ 1.96 \text{ for confidence levels of 95 percent} \\ 2.58 \text{ for confidence levels of 99 percent} \end{cases},$$

c is the confidence interval as a proportion
       (for example, 0.02 for an interval of ±2 percentage points), and

$\hat{p}$ is the expected (before measurement) proportion of households
       below the poverty line.

The scorecard, however, does not measure poverty directly, so this formula is not applicable. To derive a similar sample-size formula for the Palestine scorecard, consider the scorecard applied to the validation sample. Figure 2 shows that the expected (before measurement) poverty rate $\hat{p}$ for the national line is 0.183 (that is, the average poverty rate in the construction and calibration sub-samples). In turn, a sample size n of 16,384 and a 90-percent confidence level correspond to a confidence interval of ±0.41 percentage points (Figure 12).[14] Plugging these into the direct-measurement sample-size formula (1) above gives not $n = $ 16,384 but rather $n = \left(\dfrac{1.64}{0.0041}\right)^2 \cdot 0.183 \cdot (1 - 0.183) =$

---

[14] Due to rounding, Figure 12 displays 0.4, not 0.41.

23,922. The ratio of the sample size for scoring (derived empirically) to the sample size for direct measurement (derived from theory) is $16{,}384 \div 23{,}922 = 0.68$.

Applying the same method to $n = 8{,}192$ (confidence interval of ± 0.54 percentage points) gives $n = \left(\dfrac{1.64}{0.0054} \right)^2 \cdot 0.183 \cdot (1 - 0.183) = 13{,}791$. This time, the ratio of the sample size using scoring to the sample size using direct measurement is $8{,}192 \div 13{,}791 = 0.59$. This ratio of 0.59 for $n = 8{,}192$ is close to the ratio of 0.68 for $n = 16{,}384$. Indeed, applying this same procedure for all $n \geq 256$ in Figure 12 gives ratios that average to 0.66. This can be used to define a sample-size formula for the scorecard applied to the population in the validation sample:

$$n = \alpha \cdot \left(\frac{z}{c} \right)^2 \cdot \hat{p} \cdot (1 - \hat{p}), \tag{2}$$

where $\alpha = 0.66$ and z, c, and $\hat{p}$ are defined as in (1) above. It is this α that appears in Figure 11 under "α for sample size".

To illustrate the use of (2), suppose $c = 0.033$ (confidence interval of $\pm\, 3.3$ percentage points) and $z = 1.64$ (90-percent confidence). Then (2) gives

$n = 0.66 \cdot \left(\dfrac{1.64}{0.033} \right)^2 \cdot 0.183 \cdot (1 - 0.83) = 244$, which is close to the sample size of 256 for these parameters in Figure 12.

When the sample-size factor α is less than 1.0, it means that the scorecard is more precise than direct measurement. This occurs in all of six cases in Figure 11.

27

Of course, the sample-size formulas here are specific to Palestine, its poverty lines, its poverty rates, and this scorecard. The derivation method, however, is valid for any poverty-assessment tool following the approach in this paper.

In practice after 2005, an organization would select a poverty line (say, the national line), select a desired confidence level (say, 90 percent, or $z = 1.64$), select a desired confidence interval (say, ± 2 percentage points, or $c = 0.02$), make an assumption about $\hat{p}$ (perhaps based on a previous measurement such as the 18.4 percent national average for the 2005 PECS in Figure 2), look up α (here, 0.66 for the national line), assume that the scorecard will still work in the future,[15] and then compute the required sample size. In this illustration, $n = 0.66 \cdot \left(\dfrac{1.64}{0.02}\right)^2 0.184 \cdot (1 - 0.184)$

$= 667$.

If the scorecard has already been applied to a sample n, then $\hat{p}$ is the scorecard's estimated poverty rate, and the confidence interval c is $\pm z \cdot \sqrt{\dfrac{\alpha \cdot \hat{p} \cdot (1 - \hat{p})}{n}}$.

---

[15] This paper reports accuracy for the scorecard applied to the validation sample, but it cannot test accuracy for later years. Still, performance after the 2005 PECS will probably resemble that in the 2005 PECS, with some deterioration as time passes.

7. Estimates of changes in group poverty rates over time

The change in a group's poverty rate between two points in time is estimated as the change in the average poverty likelihood of the households in the group. With data for 2005 only, this paper cannot estimate changes over time, nor can it present sample-size formula. Nevertheless, the concepts are presented here because, in practice, pro-poor organizations can generate their own data and measure change through time.

7.1   Warning: Change is not impact

Scoring can estimate change. Of course, poverty could get better or worse, and scoring does not indicate what caused change. This point is often forgotten or confused, so it bears repeating: the scorecard simply estimates change, and it does not, in and of itself, indicate the reason for the change. In particular, estimating the impact of program participation requires knowing what would have happened to participants if they had not been participants (Moffitt, 1991). Knowing this requires either strong assumptions or a control group that resembles participants in all ways except participation. To belabor the point, the scorecard can help estimate program impact only if there is some way to know what would have happened in the absence of the program. And that information must come from somewhere beyond the scorecard. Even measuring simple change usually requires assuming that the population is constant over time and that program drop-outs do not differ from others.

7.2   Calculating estimated changes in poverty rates over time

Consider the illustration begun in the previous section. On Jan. 1, 2008, a program samples three households who score 20, 30, and 40 and so have poverty likelihoods of 53.0, 32.0, and 10.7 percent (national line, Figure 5). The group's baseline estimated poverty rate is the households' average poverty likelihood of (53.0 + 32.0 + 10.7) ÷ 3 = 31.9 percent.

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

- Score a new, independent sample, measuring change by cohort across samples
- Score the same sample at follow-up as at baseline

By way of illustration, suppose that a year later on Jan. 1, 2009, the program samples three additional households who are in the same cohort as the three households originally sampled (or suppose that the program scores the same three original households a second time) and finds that their scores are 25, 35, and 45 (poverty likelihoods of 37.2, 16.7, and 6.4 percent, national line, Figure 5). Their average poverty likelihood at follow-up is now (37.2 + 16.7 + 6.4) ÷ 3 = 20.1 percent, an improvement of 31.9 – 20.1 = 11.8 percentage points.

This suggests that about 118 of 1,000 participants crossed the poverty line in 2008.[16] Among those who started below the line, about one in three (11.8 ÷ 31.9 = 37.0 percent) ended up above the line.[17]

---

[16] This is a net figure; some people start above the line and end below it, and vice versa.
[17] The scorecard does not reveal the reasons for this change.

7.3 Accuracy for estimated change in two independent samples

With data only for 2005, it is not possible to measure the accuracy of scorecard estimates of changes in groups' poverty rates over time. In practice, of course, Palestine's scorecard can still be applied to estimate change. The following sub-sections suggest approximate sample-size formula that may be used until there is additional data.

Under direct measurement, the sample-size formula for estimates of changes in poverty rates in two equal-sized independent samples is:

$$n = 2 \cdot \left(\frac{z}{c}\right)^2 \cdot \hat{p} \cdot (1 - \hat{p}), \tag{3}$$

where z, c, and $\hat{p}$ are defined as in (1). Before measurement, $\hat{p}$ is assumed equal at both baseline and follow-up. n is the sample size at both baseline and follow-up.[18]

The method developed in the previous section can be used again to derive a sample-size formula for indirect measurement via the scorecard:

$$n = \alpha \cdot 2 \cdot \left(\frac{z}{c}\right)^2 \cdot \hat{p} \cdot (1 - \hat{p}). \tag{4}$$

As before, α is the average across sample sizes ≥ 256 of the ratio between the empirical sample size required by scoring for a given precision and the theoretical sample size required under direct measurement.

---

[18] This means that, for a given precision and with direct measurement, estimating the change in a poverty rate between two points in time requires four times as many measurements (not twice as many) as does estimating a poverty rate at a point in time.

For Peru and India (Schreiner, 2008a and 2008b), the average α across poverty lines is 1.6 and 1.2, so 1.5 may be a reasonably conservative figure for Palestine.

To illustrate the use of (4) to determine sample size for estimating changes in poverty rates across two independent samples, suppose the desired confidence level is 90 percent ($z = 1.64$), the desired confidence interval is 2 percentage points ($c = 0.02$), the poverty line is the national line, α $= 1.50$, and $\hat{p} = 0.184$ (from Figure 2). Then the baseline sample size is $n = 1.50 \cdot 2 \cdot \left(\dfrac{1.64}{0.02}\right)^2 \cdot 0.184 \cdot (1 - 0.184) = 3{,}029$, and the follow-up sample size is also 3,029.

7.4   Accuracy for estimated change for one sample, scored twice

In general, the direct-measurement sample-size formula for this case is:[19]

$$n = \left(\frac{z}{c}\right)^2 \cdot \left[\hat{p}_{12} \cdot (1 - \hat{p}_{12}) + \hat{p}_{21} \cdot (1 - \hat{p}_{21}) + 2 \cdot \hat{p}_{12} \cdot \hat{p}_{21}\right], \tag{5}$$

where z and c are defined as in (1), $\hat{p}_{12}$ is the expected (before measurement) share of all sampled cases that move from below the poverty line to above it, and $\hat{p}_{21}$ is the expected share of all sampled cases that move from above the line to below it.

How can a user set $\hat{p}_{12}$ and $\hat{p}_{21}$? Before measurement, a reasonable assumption is that the change in the poverty rate is zero. Then $\hat{p}_{12} = \hat{p}_{21} = \hat{p}_*$ and (5) becomes:

---

[19] See McNemar (1947) and Johnson (2007). John Pezzullo helped find this formula.

$$n = 2 \cdot \left(\frac{z}{c}\right)^2 \hat{p}_*.$$ (6)

Still, $\hat{p}_*$ could be anything between 0–1, so (6) is not enough to compute sample size. The estimate of $\hat{p}_*$ must be based on data available before baseline measurement.

Suppose that the observed relationship between $\hat{p}_*$ and the variance of the baseline poverty rate $p_{baseline} \cdot (1 - p_{baseline})$ is—as in Peru, see Schreiner (2008a)—close to $\hat{p}_* = 0.0085 + 0.206 \cdot [p_{baseline} \cdot (1 - p_{baseline})]$. Of course, $p_{baseline}$ is not known before baseline measurement, but it is reasonable to use as its expected value a previously observed poverty rate. Given this and a poverty line, a sample-size formula for a single sample directly measured twice for Palestine once after 2005 and then again later is:

$$n = 2 \cdot \left(\frac{z}{c}\right)^2 \cdot \{0.0085 + 0.206 \cdot [p_{2005} \cdot (1 - p_{2005})]\}.$$ (7)

As usual, (7) is multiplied by α to get scoring's sample-size formula:

$$n = \alpha \cdot 2 \cdot \left(\frac{z}{c}\right)^2 \cdot \{0.0085 + 0.206 \cdot [p_{2005} \cdot (1 - p_{2005})]\}.$$ (8)

In Peru (the only other country for which there is an estimate, Schreiner 2008a), the average α across years and poverty lines is about 1.8.

To illustrate the use of (8), suppose the desired confidence level is 90 percent (z = 1.64), the desired confidence interval is 2 percentage points ($c = 0.02$), the poverty line is the national line, and the sample will first be scored in 2006. The before-baseline poverty rate is 18.4 percent ($p_{2005} = 0.184$, Figure 2), and suppose α = 1.8. Then the

baseline sample size is $n = 1.8 \cdot 2 \cdot \left(\dfrac{1.64}{0.02}\right)^2 \cdot \{0.0085 + 0.206 \cdot [0.184 \cdot (1 - 0.184)]\} = 955$. Of

course, the same group of 954 households is scored at follow-up as well.

For a given confidence level and confidence interval, sample sizes are smaller

when one sample is scored twice than when there are two different samples.

8. Targeting

When a program uses the scorecard for targeting, households with scores at or below a cut-off are labeled *targeted* and treated—for program purposes—as if they are below a given poverty line. Households with scores above a cut-off are labeled *non-targeted* and treated—for program purposes—as if they are above a given poverty line.

There is a distinction between *targeting status* (scoring at or below a targeting cut-off) and *poverty status* (consumption below a poverty line). Poverty status is a fact that depends on whether consumption 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.

Targeting is successful when households truly below a poverty line are targeted (*inclusion*) and when households truly above a poverty line are not targeted (*exclusion*). Of course, no scorecard is perfect, and targeting is unsuccessful when households truly below a poverty line are not targeted (*undercoverage*) or when households truly above a poverty line are targeted (*leakage*). Figure 13 depicts these four possible targeting outcomes. Targeting accuracy varies by cut-off; a higher cut-off has better inclusion (but greater leakage), while a lower cut-off has better exclusion (but higher undercoverage).

A program 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

35

the four possible targeting outcomes and then to choose the cut-off that maximizes total

net benefits (Adams and Hand, 2000; Hoadley and Oliver, 1998).

Figure 14 shows the distribution of households by targeting outcome for the

scorecard applied to the validation sample. For an example cut-off of 15–19, outcomes

for the national line applied to the validation sample are:

- Inclusion: 4.3 percent are below the line and correctly targeted
- Undercoverage: 14.8 percent are below the line and mistakenly not targeted
- Leakage: 1.0 percent are above the line and mistakenly targeted
- Exclusion: 79.9 percent are above the line and correctly not targeted

Increasing the cut-off to 20–24 improves inclusion and undercoverage but

worsens leakage and exclusion:

- Inclusion: 7.6 percent are below the line and correctly targeted
- Undercoverage: 11.6 percent are below the line and mistakenly not targeted
- Leakage: 3.4 percent are above the line and mistakenly targeted
- Exclusion: 77.4 percent are above the line and correctly not targeted

Which cut-off is preferred depends on total net benefit. If each targeting outcome

has a per-household benefit or cost, then total net benefit for a given cut-off is:

Benefit per household correctly included   x   Households correctly included   +
Cost per household mistakenly not covered x   Households mistakenly not covered +
Cost per household mistakenly leaked   x   Households mistakenly leaked   +
Benefit per household correctly excluded   x   Households 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 Figure 14 for a given poverty line
- Select the cut-off with the highest total net benefit

The most difficult step is assigning benefits and costs to targeting outcomes. Any

program that uses targeting—with or without scoring—should thoughtfully consider

how it values successful inclusion or exclusion versus errors of undercoverage and

leakage. It is healthy to go through a process of thinking explicitly and intentionally

about how possible targeting outcomes are valued.

A common choice of benefits and costs is "Total Accuracy" (IRIS, 2005;

Grootaert and Braithwaite, 1998). With this, total net benefit is the number of

households correctly included or excluded:

Total Accuracy =	1	x	Households correctly included	+
	0	x	Households mistakenly undercovered	+
	0	x	Households mistakenly leaked	+
	1	x	Households correctly excluded.	

Figure 14 shows "Total Accuracy" for all cut-offs for the Palestine scorecard. For

the national line in the validation sample, total net benefit is greatest (85.0) for a cut-

off of 20–24, with more than four in five Palestinian households correctly classified.

"Total Accuracy" weighs successful inclusion of households below the line the

same as successful exclusion of households above the line. If a program valued inclusion

more (say, twice as much) than exclusion, it could reflect this by setting the benefit for

inclusion to 2 and the benefit for exclusion to 1. Then the chosen cut-off would

maximize (2 x Households correctly included) + (1 x Households correctly excluded).[20]

---

[20] Beyond "Total Accuracy", IRIS (2005) proposes a new yardstick called the "Balanced Poverty Accuracy Criterion" that is meant to account for inclusion. USAID uses BPAC as its criteria for certifying poverty-assessment tools. After normalizing by the number of people below the poverty line, the BPAC formula is:

BPAC = (Inclusion + |Undercoverage – Leakage|) x [100 ÷ (Inclusion+Undercoverage)].

As an alternative to assigning benefits and costs to targeting outcomes and then choosing a cut-off to maximize total net benefit, a program could set a cut-off to achieve a desired poverty rate among targeted households. Figure 15 shows, for the Palestine scorecard applied to the validation sample, the expected poverty rate among households who score at or below a given cut-off. For the example of the national line, targeting households who score 15–19 or less would target 5.3 percent of all Palestinian households and produce a poverty rate among those targeted of 81.8 percent.[21]

Although inclusion (and therefore targeting accuracy) is in the BPAC formula, BPAC is maximized by minimizing the difference between undercoverage and leakage, regardless of inclusion. But the difference between undercoverage and leakage is the same as the difference between the estimated poverty rate and the true poverty rate. Thus, it would be clearer to discard the BPAC nomenclature and speak directly in terms of the accuracy of the estimated poverty rate.

[21] If potential participants are not representative of all of Palestine, then Figure 15 is valid only if selection into potential participation—whether by the program or potential participant—is unrelated with poverty in any way not captured by the scorecard.

9. Conclusion

This paper presents the scorecard. Pro-poor programs in Palestine (West Bank and Gaza Strip) can use it to estimate the likelihood that a household has consumption below a given poverty line, to estimate the poverty rate of a group of households at a point in time, and to estimate changes in the poverty rate of a group of households between two points in time. The scorecard can also be used for targeting.

The scorecard is inexpensive to use and can be understood by non-specialists. It is designed to be practical for local pro-poor organizations who want to improve how they monitor and manage their social performance in order to speed up their participants' progress out of poverty.

The scorecard is built with a sub-sample of data from the 2005 PECS, tested with a different sub-sample, and calibrated to six poverty lines (national, "subsistence", USAID "extreme", $1/day, $2/day, and $3/day).

Accuracy and sample-size formulas are reported for estimates of households' poverty likelihoods, groups' poverty rates at a point in time, and changes in groups' poverty rates over time. Of course, the scorecard's estimates of changes in poverty rates are not the same as estimates of program impact.

When the scorecard is applied to the validation sample, the difference between estimates versus true poverty rates for groups of households at a point in time is always less than 2.4 percentage points and averages—across the six poverty lines—about 1.0 percentage points. For $n = 16,384$ and 90-percent confidence, the precision of these

differences is ± 0.5 percentage points or less, and for $n = 1{,}024$, precision is ± 2.0 percentage points or less.

For targeting, programs can use the results reported here to select a cut-off that fits their values and mission.

Although the statistical technique is innovative, and although technical accuracy is important, the design of the scorecard here focuses on transparency and ease-of-use. After all, a perfectly accurate scorecard is worthless if programs feel so daunted by its complexity or its cost that they do not even try to use it. For this reason, the scorecard is kept simple, using 10 indicators that are inexpensive to collect and that are straightforward to verify. Points are all zeros or positive integers, and scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line). Scores are related to poverty likelihoods via simple look-up tables, and targeting cut-offs are likewise simple to apply. The design attempts to facilitate adoption by helping managers understand and trust scoring and by allowing non-specialists to generate scores quickly in the field.

In sum, the scorecard is a practical, objective way for pro-poor programs in Palestine to monitor poverty rates, track changes in poverty rates over time, and target services. The same approach can be applied to any country with similar data from a national consumption survey.

References

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

Baesens, B.; Van Gestel, T.; Viaene, S.; Stepanova, M.; Suykens, J.; and J. Vanthienen. (2003) "Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring", *Journal of the Operational Research Society*, Vol. 54, pp. 627–635.

Caire, Dean. (2004) "Building Credit Scorecards for Small Business Lending in Developing Markets", `microfinance.com/English/Papers/Scoring_SMEs_Hybrid.pdf`, accessed September 9, 2008.

Central Bureau of Statistics. (2006) "The Palestinian Expenditure and Consumption Survey 2005: Levels of Living in the Palestinian Territory, Final Report (January 2005 – January 2006)", `www.pcbs.gov.ps/Portals/_pcbs/living/report_english.pdf`, accessed December 9, 2008.

Coady, David; Grosh, Margaret; and John Hoddinott. (2004) *Targeting of Transfers in Developing Countries*, `hdl.handle.net/10986/14902`, retrieved 13 May 2016.

Cochran, William G. (1977) *Sampling Techniques, Third Edition*.

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

Deaton, Angus; and Salman Zaidi. (2002) "Guidelines for Constructing Consumption Aggregates for Welfare Analysis", World Bank LSMS Working Paper No. 135, `princeton.edu/rpds/papers/pdfs/deaton_zaidi_consumption.pdf`, accessed December 5, 2008.

Efron, Bradley; and Robert J. Tibshirani. (1993) *An Introduction to the Bootstrap*.

Friedman, Jerome H. (1997) "On Bias, Variance, 0–1 Loss, and the Curse-of-Dimensionality", *Data Mining and Knowledge Discovery*, Vol. 1, pp. 55–77.

Fuller, Rob. (2006) "Measuring Poverty of Microfinance Clients in Haiti", `microfinance.com/English/Papers/Scoring_Poverty_Haiti_Fuller.pdf`, accessed September 9, 2008.

Goodman, L.A.; and Kruskal, W.H. (1979) *Measures of Association for Cross Classification*.

Grootaert, Christiaan; and Jeanine Braithwaite. (1998) "Poverty Correlates and Indicator-Based Targeting in Eastern Europe and the Former Soviet Union", World Bank Policy Research Working Paper No. 1942, `dx.doi.org/10.1596/1813-9450-1942`, retrieved 15 May 2016.

Grosh, Margaret; and Judy L. Baker. (1995) "Proxy-Means Tests for Targeting Social Programs: Simulations and Speculation", World Bank LSMS Working Paper No. 118, `go.worldbank.org/W9OWN57PD0`, retrieved 13 May 2016.

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

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.

IRIS Center. (2007a) "Manual for the Implementation of USAID Poverty Assessment Tools", `povertytools.org/training_documents/Manuals/USAID_PAT_Manual_Eng.pdf`, accessed September 9, 2008.

_____. (2007b) "Introduction to Sampling for the Implementation of PATs", `povertytools.org/training_documents/Sampling/Introduction_Sampling.ppt`, accessed September 9, 2008.

_____. (2005) "Notes on Assessment and Improvement of Tool Accuracy", `povertytools.org/other_documents/AssessingImproving_Accuracy.pdf`, accessed September 9, 2008.

Johnson, Glenn. (2007) "Lesson 3: Two-Way Tables—Dependent Samples", `www.stat.psu.edu/online/development/stat504/03_2way/53_2way_compare.htm`, accessed September 9, 2008.

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

Lovie, A.D.; and P. Lovie. (1986) "The Flat Maximum Effect and Linear Scoring Models for Prediction", *Journal of Forecasting*, Vol. 5, pp. 159–168.

Matul, Michal; and Sean Kline. (2003) "Scoring Change: Prizma's Approach to Assessing Poverty", Microfinance Centre for Central and Eastern Europe and the New Independent States Spotlight Note No. 4, `www.mfc.org.pl/doc/Research/ImpAct/SN/MFC_SN04_eng.pdf`, accessed September 9, 2008.

McNemar, Quinn. (1947) "Note on the Sampling Error of the Difference between Correlated Proportions or Percentages", *Psychometrika*, Vol. 17, pp. 153–157.

Moffitt, Robert. (1991) "Program Evaluation with Non-experimental Data", *Evaluation Review*, Vol. 15, No. 3, pp. 291–314.

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.

Narayan, Ambar; and Nobuo Yoshida. (2005) "Proxy Means Tests for Targeting Welfare Benefits in Sri Lanka", World Bank Report No. SASPR–7, `documents.worldbank.org/curated/en/2005/07/6209268/proxy-means-test-targeting-welfare-benefits-sri-lanka`, retrieved 5 May 2016.

SAS Institute Inc. (2004) "The LOGISTIC Procedure: Rank Correlation of Observed Responses and Predicted Probabilities", in *SAS/STAT User's Guide, Version 9.*

Schreiner, Mark. (2013) "Simple Poverty Scorecard Poverty-Assessment Tool: Bangladesh", `SimplePovertyScorecard.com/BGD_2010_ENG.pdf`, accessed 6 July 2016.

_____. (2008a) "Simple Poverty Scorecard Poverty-Assessment Tool: Peru", `SimplePovertyScorecard.com/PER_2003_ENG.pdf`, accessed 6 July 2016.

_____. (2008b) "Simple Poverty Scorecard Poverty-Assessment Tool: India", `SimplePovertyScorecard.com/IND_2005_ENG.pdf`, accessed 6 July 2016.

_____. (2006) "Is One Simple Poverty Scorecard Poverty-Assessment Tool Enough for India?", `microfinance.com/English/Papers/Scoring_Poverty_India_Segments.pdf`, accessed September 9, 2008.

_____. (2005a) "La Herramienta del Índice de Calificación de la Pobreza™: México", `SimplePovertyScorecard.com/MEX_2002_SPA.pdf`, accessed 6 July 2016.

_____. (2005b) "IRIS Questions on the Simple Poverty Scorecard Poverty-Assessment Tool", `microfinance.com/English/Papers/Scoring_Poverty_Response_to_IRIS.pdf`, accessed September 9, 2008.

_____. (2002) *Scoring: The Next Breakthrough in Microfinance?* CGAP Occasional Paper No. 7, `microfinance.com/English/Papers/Scoring_Breakthrough_CGAP.pdf`, retrieved 13 May 2016.

_____; Matul, Michal; Pawlak, Ewa; and Sean Kline. (2004) "Poverty Scoring: Lessons from a Microlender in Bosnia-Herzegovina", `microfinance.com/English/Papers/Scoring_Poverty_in_BiH_Short.pdf`, accessed September 9, 2008.

Sillers, Don. (2006) "National and International Poverty Lines: An Overview", `pdf.usaid.gov/pdf_docs/Pnadh069.pdf`, retrieved 13 May 2016.

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.

Toohig, Jeff. (2007) "Progress Out of Poverty Index®: Training Guide", `progressoutofpoverty.org/toolkit`, accessed September 9, 2008.

United States Congress. (2004) "Microenterprise Results and Accountability Act of 2004 (HR 3818 RDS)", November 20, `smith4nj.com/laws/108-484.pdf`, retrieved 13 May 2016.

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

World Bank. (2004) "Four Years—Intifada, Closures and Palestinian Economic Crisis: An Assessment", `siteresources.worldbank.org/INTWESTBANKGAZA/Resources/wbgaza-4yrassessment.pdf`, accessed September 9, 2008.

Zeller, Manfred. (2004) "Review of Poverty Assessment Tools", `pdf.usaid.gov/pdf_docs/PNADH120.pdf`, retrieved 13 May 2016.

Figure 2: Sample sizes and household poverty rates by sub-sample and poverty line

		% with expenditure below a poverty line					
			National	USAID		International	
Sub-sample	Households	National	'Subsistence'	'Extreme'	$1/day	$2/day	$3/day
All Palestine	2,152	18.4	6.1	9.1	2.5	20.7	43.9
Construction							
Selecting indicators and weights	708	18.2	7.0	10.5	2.1	21.0	43.6
Calibration							
Associating scores with likelihoods	752	18.5	5.8	8.3	2.5	20.6	43.9
Validation							
Applying scorecards	692	18.4	5.6	8.4	2.9	20.6	44.0
Change between construction and calibration to validation (percentage points)							
		-0.1	0.8	1.0	-0.5	0.2	-0.3

Source: 2005 PECS

Note: For the accuracy tests, nine households in the calibration sample with scores of nine or less are put into the validation sample to ensure of all subsamples have some cases in all score ranges. This barely affects the results.

Figure 3: Average poverty lines and poverty rates by region (household level)

| Area | Line or rate | Poverty line (NIS/person/day) and poverty rate (%) | | | | | |
		National	National 'Subsistence'	USAID 'Extreme'	$1/day	International $2/day	$3/day
West Bank	Line	10.94	7.48	8.28	5.76	11.52	17.28
	Rate	14.5	4.8	7.2	1.8	16.4	37.4
Gaza Strip	Line	9.78	6.68	7.63	5.15	10.30	15.45
	Rate	26.1	8.7	12.7	3.9	29.3	56.5
All Palestine:	Line	10.55	7.21	8.06	5.55	11.11	16.66
	Rate	18.4	6.1	9.1	2.5	20.7	43.9

Source: 2005 PECS

Figure 3: Average poverty lines and poverty rates by region (person level)

| Area | Line or rate | Poverty line (NIS/person/day) and poverty rate (%) | | | | | |
		National	National 'Subsistence'	USAID 'Extreme'	$1/day	International $2/day	$3/day
West Bank	Line	10.94	7.48	8.28	5.76	11.52	17.28
	Rate	18.6	7.0	10.0	2.7	20.9	45.5
Gaza Strip	Line	9.78	6.68	7.63	5.15	10.30	15.45
	Rate	31.7	10.7	15.3	4.9	35.3	64.5
All Palestine:	Line	10.49	7.17	8.03	5.52	11.05	16.57
	Rate	23.6	8.4	12.0	3.5	26.5	52.8

Source: 2005 PECS

Figure 4: Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly indicative of poverty)
153	How many household members are 18-years-old or younger? (Six or more; Five; Three or four; Two; One; None)
115	How many female household members are 18-years-old or younger? (Five or more; Four; Three; Two; One; None)
86	How many male household members are 18-years-old or younger? (Five or more; Four; Three; Two; One; None)
76	How many children ages 6 to 12 attend school? (Not all; All; No children in the age range)
61	Does the household have a phone line? (No; Yes)
60	How many male children ages 6 to 17 attend school? (Not all; All; No children in the age range)
58	What is the household's main source of energy for heating? (Coal/firewood, electricity, none, or no data; Gas, kerosene, or other)
57	Does the household have a private car? (No; Yes)
46	What is the highest education level passed by any household member? (None, elementary, or preparatory education; Secondary education; Associate diploma, bachelor's degree, or higher)
41	Does the household have a satellite dish? (No; Yes)
35	Does the household have an electric sweeper? (No; Yes)
32	How old is the male head/spouse? (31 or younger; 32 to 35; 46 or older)
31	Does the household have a solar boiler? (No; Yes)
30	Does the household have access to the internet? (No; Yes)
29	Does the household have a computer? (No; Yes)
28	What is the highest education level passed by the male head/spouse? (None, elementary, or preparatory education; Secondary education, associate diploma, bachelor's degree or higher)
28	Does the household have a video player? (No; Yes)
26	Does the household have a home library? (No; Yes)

Source: 2005 PECS, national poverty line.

Figure 4 (continued): Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly associated with poverty)
25	How many household members ages 6 to 25 attend school? (Not all; All; No children in the age range)
21	Does the household have a mobile phone? (No; Yes)
19	What is the household's main source of income? (Subsistence farming, household business, wages and salaries from private sector, or social affairs; Wages and salaries from public sector or Israeli sector, or cash remittances received from within Palestine; Cash remittances received from abroad, international organizations, or other)
19	What sector does the male head/spouse work in? (International organizations, or outside establishments; Foreign private sectors; National private sectors; National or foreign governmental sectors, non-profit organizations, or no male head)
16	What is the female head's employment status? (Self-employed, or other; Employed, employer, or no female head)
15	Does any household member work in outside establishments? (No; Yes)
14	What sector does the female head/spouse work in? (Outside establishments; National/foreign private sectors, national/foreign governmental sectors, non-profit organizations, international organizations, or no female head)
13	What is the male head's employment status? (Self-employed; Employed or not applicable; No male head; Employer)
13	What type of accommodation does the household have? (House; Villa, apartment, or other (room, tent, marginal, etc.))
12	Is any household member self-employed? (No; Yes)
11	Does the female head/spouse work? (No; Yes)
10	Does the household have agricultural land (be it owned, free, or rented)? (No; Yes)
10	What is the structure of household headship? (Female head/spouse only; Male head/spouse only; Both male and female heads/spouses)
10	What type of connection to sewage networks does the household have? (Cesspit, or no connection; Public sewage system)
8	Does the household live in a refugee camp? (No; Yes)
8	How many household members work? (None, one, or two; Three or more)
8	Does the household have a refrigerator? (No; Yes)

Source: 2005 PECS, national poverty line.

National Poverty Line Tables

(and tables pertaining to all six poverty lines)

Figure 5 (National poverty line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	100.0
5–9	100.0
10–14	83.0
15–19	53.8
20–24	53.0
25–29	37.2
30–34	32.0
35–39	16.7
40–44	10.7
45–49	6.4
50–54	7.1
55–59	4.4
60–64	4.7
65–69	0.0
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Surveyed cases weighted to represent Palestine's households.

Based on the 2005 PECS.

Figure 6 (National poverty line): Derivation of estimated poverty likelihoods associated with scores

Score	Households below poverty line		All households at score		Poverty likelihood (estimated, %)
0–4	422	÷	422	=	100.0
5–9	782	÷	782	=	100.0
10–14	1,642	÷	1,978	=	83.0
15–19	1,134	÷	2,105	=	53.8
20–24	3,036	÷	5,726	=	53.0
25–29	3,720	÷	9,999	=	37.2
30–34	2,927	÷	9,156	=	32.0
35–39	1,483	÷	8,897	=	16.7
40–44	1,465	÷	13,652	=	10.7
45–49	568	÷	8,888	=	6.4
50–54	740	÷	10,402	=	7.1
55–59	396	÷	8,908	=	4.4
60–64	259	÷	5,481	=	4.7
65–69	0	÷	4,748	=	0.0
70–74	0	÷	4,446	=	0.0
75–79	0	÷	1,987	=	0.0
80–84	0	÷	856	=	0.0
85–89	0	÷	851	=	0.0
90–94	0	÷	392	=	0.0
95–100	0	÷	322	=	0.0

Surveyed cases weighted to represent Palestine's households.

Based on the 2005 PECS.

Figure 7 (All poverty lines): Distribution of household poverty likelihoods across ranges demarcated by poverty lines

	Likelihood of having expenditure in range demarcated by daily per capita poverty lines						
	<$1/day	≥$1/day and <'Subsistence'	≥'Subsistence' and <USAID	≥USAID and <National	≥National and <$2/day	≥$2/day and <$3/day	≥$3/day
Score	<NIS5.55	≥NIS5.55 and <NIS7.21	≥NIS7.21 and <NIS8.06	≥NIS8.06 and <NIS10.55	≥NIS10.55 and <NIS11.11	≥NIS11.11 and <NIS16.66	≥NIS16.66
0–4	56.0	44.0	0.0	0.0	0.0	0.0	0.0
5–9	51.8	8.3	0.0	39.9	0.0	0.0	0.0
10–14	22.4	24.6	8.5	27.6	9.4	0.0	7.6
15–19	15.3	12.8	9.9	15.8	16.0	22.4	7.8
20–24	4.0	16.0	8.5	24.6	1.8	38.0	7.1
25–29	5.0	5.7	3.6	23.0	3.5	35.1	24.2
30–34	3.6	3.4	5.4	19.6	5.2	37.0	25.8
35–39	0.0	2.0	2.0	12.6	2.8	35.8	44.8
40–44	1.0	1.2	2.4	6.1	0.0	33.1	56.2
45–49	0.0	0.0	0.0	6.4	0.0	28.2	65.4
50–54	0.0	0.0	2.7	4.4	1.2	11.5	80.2
55–59	0.0	2.2	0.0	2.3	0.0	6.2	89.4
60–64	0.0	0.0	0.0	4.7	0.0	9.7	85.6
65–69	0.0	0.0	0.0	0.0	0.0	7.1	92.9
70–74	0.0	0.0	0.0	0.0	0.0	4.3	95.7
75–79	0.0	0.0	0.0	0.0	0.0	5.8	94.2
80–84	0.0	0.0	0.0	0.0	0.0	0.0	100.0
85–89	0.0	0.0	0.0	0.0	0.0	0.0	100.0
90–94	0.0	0.0	0.0	0.0	0.0	0.0	100.0
95–100	0.0	0.0	0.0	0.0	0.0	0.0	100.0

All poverty likelihoods in percentage units.

Figure 8 (National poverty line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample ($n = 16{,}384$) from the validation sample, with confidence intervals

| | | Difference between estimate and true value | | |
| | | Confidence interval ($+/-$ percentage points) | | |
Score	Diff.	90-percent	95-percent	99-percent
0–4	0.0	0.0	0.0	0.0
5–9	0.0	0.0	0.0	0.0
10–14	0.7	3.7	4.4	5.9
15–19	-10.5	7.4	7.8	8.8
20–24	-2.2	2.7	3.2	3.9
25–29	-10.6	6.4	6.5	6.8
30–34	-4.0	3.0	3.2	3.6
35–39	1.8	1.5	1.8	2.4
40–44	4.3	0.9	1.1	1.4
45–49	1.4	0.9	1.1	1.4
50–54	4.8	0.6	0.7	0.9
55–59	1.6	0.7	0.8	1.1
60–64	4.7	0.0	0.0	0.0
65–69	0.0	0.0	0.0	0.0
70–74	-2.9	2.0	2.1	2.3
75–79	0.0	0.0	0.0	0.0
80–84	0.0	0.0	0.0	0.0
85–89	0.0	0.0	0.0	0.0
90–94	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0

Based on scorecard applied to the validation sample.

Figure 9 (All poverty lines): Differences and precision of differences for bootstrapped estimates of households' poverty likelihoods for the validation sample

	Poverty line					
	National	National 'Subsistence'	USAID 'Extreme'	$1/day	$2/day	$3/day
Estimate minus true value	-0.6	-1.4	-1.4	0.4	-0.6	-0.9
Precision of difference	0.4	0.5	0.5	0.7	0.3	0.3

Precision is measured as 90-percent confidence intervals in units of +/− percentage points.

Differences and precision estimated from 1,000 bootstraps of size $n = 16{,}384$.

Figure 10 (National poverty line): Differences and precision of differences for bootstrapped estimates of households' poverty likelihoods, by sample size, scorecard applied to validation sample

| Sample size (n) | Diff. | Difference between estimate and true value | | |
| | | Confidence interval (+/− percentage points) | | |
		90-percent	95-percent	99-percent
2	-0.7	36.1	43.7	62.0
4	-0.5	26.3	32.1	44.8
8	-0.5	19.7	23.3	31.2
16	-0.6	14.5	17.4	22.2
32	-0.5	10.7	12.8	16.7
64	-0.6	7.4	8.8	12.8
128	-0.7	5.3	6.3	8.8
256	-0.7	3.4	4.2	5.6
512	-0.6	2.3	2.7	3.5
1,024	-0.6	1.5	1.7	2.3
2,048	-0.6	1.0	1.3	1.7
4,096	-0.6	0.7	0.9	1.1
8,192	-0.6	0.5	0.6	0.8
16,384	-0.6	0.4	0.4	0.6

Figure 11 (All poverty lines): Differences, precision of differences, and sample-size α for bootstrapped estimates of poverty rates for groups of households at a point in time for the scorecard applied to the validation sample

	Poverty line					
	National	National 'Subsistence'	USAID 'Extreme'	$1/day	$2/day	$3/day
Estimate minus true value	-0.1	-1.6	-2.4	-0.1	-1.1	-0.7
Precision of difference	0.4	0.3	0.3	0.2	0.4	0.5
α for sample size	0.66	0.99	0.94	0.88	0.67	0.68

Precision is measured as 90-percent confidence intervals in units of $+/-$ percentage points.

Differences and precision estimated from 1,000 bootstraps of size $n = 16{,}384$.

α is estimated from 1,000 bootstrap samples of $n = 256, 512, 1{,}024, 2{,}048, 4{,}096, 8{,}192$, and 16,384.

Figure 12 (National poverty line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to validation sample

| Sample size (n) | Diff. | Difference between estimate and true value Confidence interval (+/− percentage points) | | |
		90-percent	95-percent	99-percent
2	-0.7	36.1	43.7	62.0
4	-0.5	26.1	32.3	44.8
8	-0.5	18.4	21.6	28.9
16	-0.3	13.0	16.0	20.1
32	-0.2	9.5	12.1	14.4
64	-0.2	6.2	7.5	10.1
128	-0.3	4.6	5.5	7.2
256	-0.2	3.3	3.9	5.1
512	-0.2	2.3	2.8	3.7
1,024	-0.1	1.6	1.9	2.6
2,048	-0.1	1.2	1.4	1.8
4,096	-0.1	0.8	1.0	1.2
8,192	-0.1	0.5	0.7	0.9
16,384	-0.1	0.4	0.5	0.6

Figure 13 (All poverty lines): Possible types of outcomes from targeting by poverty score

		Targeting segment	
		Targeted	**Non-targeted**
True poverty status	**Below poverty line**	**Inclusion** Under poverty line Correctly targeted	**Undercoverage** Under poverty line Mistakenly non-targeted
	Above poverty line	**Leakage** Above poverty line Mistakenly targeted	**Exclusion** Above poverty line Correctly non-targeted

Figure 14 (National poverty line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: ≥ poverty line mistakenly targeted	Exclusion: ≥ poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.4	18.7	0.0	80.8	81.3	-95.6
5–9	1.2	18.0	0.0	80.8	82.0	-87.4
10–14	2.9	16.3	0.3	80.5	83.4	-68.4
15–19	4.3	14.8	1.0	79.9	84.2	-49.8
20–24	7.6	11.6	3.4	77.4	85.0	-2.9
25–29	12.5	6.7	8.5	72.3	84.8	55.6
30–34	15.8	3.3	14.3	66.5	82.3	25.3
35–39	17.2	2.0	21.9	58.9	76.1	-14.3
40–44	18.1	1.1	34.7	46.2	64.2	-80.8
45–49	18.5	0.7	43.1	37.7	56.2	-124.9
50–54	18.8	0.4	53.2	27.6	46.3	-177.8
55–59	19.0	0.1	61.9	18.9	38.0	-222.9
60–64	19.0	0.1	67.4	13.5	32.5	-251.5
65–69	19.0	0.1	72.1	8.7	27.7	-276.2
70–74	19.2	0.0	76.4	4.4	23.6	-298.7
75–79	19.2	0.0	78.4	2.4	21.6	-309.1
80–84	19.2	0.0	79.3	1.6	20.7	-313.5
85–89	19.2	0.0	80.1	0.7	19.9	-318.0
90–94	19.2	0.0	80.5	0.3	19.5	-320.0
95–100	19.2	0.0	80.8	0.0	19.2	-321.7

Inclusion, undercoverage, leakage, and exclusion normalized to sum to 100.

Figure 15 (National poverty line): Households below the poverty line and all households, at a given score or at or below a given score cut-off, scorecard applied to validation sample

Score	Households below poverty line (%)		All households (%)	
	At score	At or below score	At score	At or below score
0–4	100.0	100.0	0.4	0.4
5–9	100.0	100.0	0.8	1.2
10–14	84.4	90.3	2.0	3.2
15–19	68.9	81.8	2.1	5.3
20–24	57.3	69.1	5.7	11.0
25–29	48.9	59.5	10.0	21.0
30–34	36.5	52.5	9.2	30.2
35–39	14.8	43.9	8.9	39.1
40–44	6.6	34.3	13.7	52.7
45–49	4.9	30.0	8.9	61.6
50–54	2.5	26.1	10.4	72.0
55–59	3.0	23.5	8.9	80.9
60–64	0.0	22.0	5.5	86.4
65–69	0.0	20.9	4.7	91.1
70–74	3.1	20.1	4.4	95.6
75–79	0.0	19.6	2.0	97.6
80–84	0.0	19.5	0.9	98.4
85–89	0.0	19.3	0.9	99.3
90–94	0.0	19.2	0.4	99.7
95–100	0.0	19.2	0.3	100.0

Figure 16: Poverty rates by region for "consumption" and "expenditure" definitions in the 2005 PECS

Area	Poverty rate (%, household level)				Poverty rate (%, person level)			
	National line		National "Subsistence" line		National line		National "Subsistence" line	
	Consumption	Expenditure	Consumption	Expenditure	Consumption	Expenditure	Consumption	Expenditure
West Bank	14.5	22.7	4.8	10.8	18.6	27.7	7.0	13.8
Gaza Strip	26.1	34.8	8.7	15.7	31.7	39.5	10.7	18.3
All Palestine:	18.4	26.8	6.1	12.5	23.6	32.2	8.4	15.5

Source: 2005 PECS

National "Subsistence" Poverty Line Tables

Figure 5 (National "subsistence" line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	100.0
5–9	60.2
10–14	47.0
15–19	28.2
20–24	19.9
25–29	10.7
30–34	7.0
35–39	2.0
40–44	2.2
45–49	0.0
50–54	0.0
55–59	2.2
60–64	0.0
65–69	0.0
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Surveyed cases weighted to represent Palestine's households.

Based on the 2005 PECS.

Figure 6 (National "subsistence" line): Derivation of estimated poverty likelihoods associated with scores

Score	Households below poverty line		All households at score		Poverty likelihood (estimated, %)
0–4	422	÷	422	=	100.0
5–9	470	÷	782	=	60.2
10–14	929	÷	1,978	=	47.0
15–19	593	÷	2,105	=	28.2
20–24	1,141	÷	5,726	=	19.9
25–29	1,065	÷	9,999	=	10.7
30–34	645	÷	9,156	=	7.0
35–39	182	÷	8,897	=	2.0
40–44	300	÷	13,652	=	2.2
45–49	0	÷	8,888	=	0.0
50–54	0	÷	10,402	=	0.0
55–59	193	÷	8,908	=	2.2
60–64	0	÷	5,481	=	0.0
65–69	0	÷	4,748	=	0.0
70–74	0	÷	4,446	=	0.0
75–79	0	÷	1,987	=	0.0
80–84	0	÷	856	=	0.0
85–89	0	÷	851	=	0.0
90–94	0	÷	392	=	0.0
95–100	0	÷	322	=	0.0

Surveyed cases weighted to represent Palestine's households.

Based on the 2005 PECS.

Figure 8 (National "subsistence" line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample ($n = 16{,}384$) from the validation sample, with confidence intervals

Score	Diff.	Difference between estimate and true value Confidence interval (+/− percentage points) 90-percent	95-percent	99-percent
0–4	0.0	0.0	0.0	0.0
5–9	5.3	7.7	9.1	11.4
10–14	13.3	4.3	5.0	6.5
15–19	-29.1	16.8	17.2	17.7
20–24	-3.2	2.7	2.9	3.7
25–29	-2.3	1.8	1.9	2.1
30–34	-9.5	5.6	5.7	6.0
35–39	-2.3	1.6	1.7	1.8
40–44	1.4	0.3	0.3	0.5
45–49	0.0	0.0	0.0	0.0
50–54	0.0	0.0	0.0	0.0
55–59	2.2	0.0	0.0	0.0
60–64	0.0	0.0	0.0	0.0
65–69	0.0	0.0	0.0	0.0
70–74	-2.9	2.0	2.1	2.3
75–79	0.0	0.0	0.0	0.0
80–84	0.0	0.0	0.0	0.0
85–89	0.0	0.0	0.0	0.0
90–94	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0

Based on scorecard applied to the validation sample.

Figure 10 (National "subsistence" line): Differences and precision of differences for bootstrapped estimates of households' poverty likelihoods, by sample size, scorecard applied to validation sample

Sample size (n)	Diff.	Difference between estimate and true value		
		Confidence interval (+/− percentage points)		
		90-percent	95-percent	99-percent
2	-1.9	30.6	36.1	56.1
4	-1.5	20.0	25.0	34.0
8	-1.5	14.6	18.3	23.2
16	-1.5	11.4	13.4	18.5
32	-1.7	9.4	11.2	14.3
64	-1.7	7.3	8.3	11.1
128	-1.8	6.0	7.4	9.4
256	-1.7	4.4	5.2	6.9
512	-1.5	3.2	3.9	4.9
1,024	-1.4	2.0	2.4	3.4
2,048	-1.4	1.4	1.7	2.1
4,096	-1.4	1.0	1.2	1.6
8,192	-1.4	0.7	0.8	1.2
16,384	-1.4	0.5	0.6	0.8

Figure 12 (National "subsistence" line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to validation sample

| Sample size (n) | Diff. | Difference between estimate and true value | | |
| | | Confidence interval (+/− percentage points) | | |
		90-percent	95-percent	99-percent
2	-1.9	30.6	36.1	56.1
4	-1.6	17.9	24.7	32.4
8	-1.6	13.9	15.8	21.9
16	-1.5	9.3	11.3	14.9
32	-1.5	6.7	8.2	10.9
64	-1.5	4.5	5.6	7.4
128	-1.6	3.3	4.1	5.3
256	-1.6	2.3	2.7	3.7
512	-1.6	1.7	2.0	2.7
1,024	-1.5	1.2	1.4	1.8
2,048	-1.5	0.9	1.0	1.4
4,096	-1.5	0.6	0.7	0.9
8,192	-1.5	0.4	0.5	0.6
16,384	-1.6	0.3	0.3	0.5

Figure 14 (National "subsistence" line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: ≥ poverty line mistakenly targeted	Exclusion: ≥ poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.4	7.4	0.0	92.2	92.6	-89.2
5–9	0.9	7.0	0.3	91.8	92.7	-73.3
10–14	1.6	6.2	1.6	90.6	92.2	-38.7
15–19	2.9	5.0	2.4	89.8	92.6	4.2
20–24	4.3	3.6	6.8	85.4	89.6	13.8
25–29	5.6	2.2	15.4	76.8	82.4	-96.1
30–34	7.2	0.7	23.0	69.2	76.3	-193.2
35–39	7.6	0.3	31.5	60.7	68.2	-301.4
40–44	7.7	0.1	45.0	47.1	54.8	-473.8
45–49	7.7	0.1	53.9	38.3	46.0	-587.1
50–54	7.7	0.1	64.3	27.9	35.6	-719.7
55–59	7.7	0.1	73.2	18.9	26.7	-833.3
60–64	7.7	0.1	78.7	13.5	21.2	-903.2
65–69	7.7	0.1	83.4	8.7	16.4	-963.7
70–74	7.8	0.0	87.7	4.4	12.3	-1,018.6
75–79	7.8	0.0	89.7	2.4	10.3	-1,043.9
80–84	7.8	0.0	90.6	1.6	9.4	-1,054.8
85–89	7.8	0.0	91.4	0.7	8.6	-1,065.7
90–94	7.8	0.0	91.8	0.3	8.2	-1,070.7
95–100	7.8	0.0	92.2	0.0	7.8	-1,074.8

Inclusion, undercoverage, leakage, and exclusion normalized to sum to 100.

Figure 15 (National "subsistence" line): Households below the poverty line and all households, at a given score or at or below a given score cut-off, scorecard applied to validation sample

Score	Households below poverty line (%)		All households (%)	
	At score	At or below score	At score	At or below score
0–4	100.0	100.0	0.4	0.4
5–9	60.1	74.1	0.8	1.2
10–14	37.3	51.2	2.0	3.2
15–19	59.7	54.6	2.1	5.3
20–24	23.9	38.6	5.7	11.0
25–29	13.8	26.8	10.0	21.0
30–34	16.8	23.8	9.2	30.2
35–39	4.6	19.4	8.9	39.1
40–44	0.9	14.6	13.7	52.7
45–49	0.0	12.5	8.9	61.6
50–54	0.0	10.7	10.4	72.0
55–59	0.0	9.5	8.9	80.9
60–64	0.0	8.9	5.5	86.4
65–69	0.0	8.5	4.7	91.1
70–74	3.1	8.2	4.4	95.6
75–79	0.0	8.0	2.0	97.6
80–84	0.0	8.0	0.9	98.4
85–89	0.0	7.9	0.9	99.3
90–94	0.0	7.9	0.4	99.7
95–100	0.0	7.8	0.3	100.0

USAID "Extreme" Poverty Line Tables

Figure 5 (USAID "extreme" line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	100.0
5–9	60.2
10–14	55.4
15–19	38.1
20–24	28.5
25–29	14.2
30–34	12.4
35–39	4.1
40–44	4.6
45–49	0.0
50–54	2.7
55–59	2.2
60–64	0.0
65–69	0.0
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Surveyed cases weighted to represent Palestine's households.

Based on the 2005 PECS.

Figure 6 (USAID "extreme" line): Derivation of estimated poverty likelihoods associated with scores

Score	Households below poverty line		All households at score		Poverty likelihood (estimated, %)
0–4	422	÷	422	=	100.0
5–9	470	÷	782	=	60.2
10–14	1,096	÷	1,978	=	55.4
15–19	801	÷	2,105	=	38.1
20–24	1,630	÷	5,726	=	28.5
25–29	1,420	÷	9,999	=	14.2
30–34	1,134	÷	9,156	=	12.4
35–39	361	÷	8,897	=	4.1
40–44	629	÷	13,652	=	4.6
45–49	0	÷	8,888	=	0.0
50–54	279	÷	10,402	=	2.7
55–59	193	÷	8,908	=	2.2
60–64	0	÷	5,481	=	0.0
65–69	0	÷	4,748	=	0.0
70–74	0	÷	4,446	=	0.0
75–79	0	÷	1,987	=	0.0
80–84	0	÷	856	=	0.0
85–89	0	÷	851	=	0.0
90–94	0	÷	392	=	0.0
95–100	0	÷	322	=	0.0

Surveyed cases weighted to represent Palestine's households.

Based on the 2005 PECS.

Figure 8 (USAID "extreme" line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample ($n = 16{,}384$) from the validation sample, with confidence intervals

Score	Diff.	Difference between estimate and true value		
		Confidence interval ($+/-$ percentage points)		
		90-percent	95-percent	99-percent
0–4	0.0	0.0	0.0	0.0
5–9	5.3	7.7	9.1	11.4
10–14	21.7	4.3	5.0	6.5
15–19	-25.3	14.8	15.2	16.1
20–24	-5.5	4.0	4.2	4.8
25–29	-8.1	4.9	5.0	5.4
30–34	-16.7	9.3	9.6	10.0
35–39	-0.2	0.8	1.0	1.3
40–44	1.9	0.6	0.7	0.9
45–49	-1.2	0.8	0.9	1.0
50–54	1.0	0.5	0.6	0.9
55–59	2.2	0.0	0.0	0.0
60–64	0.0	0.0	0.0	0.0
65–69	0.0	0.0	0.0	0.0
70–74	-2.9	2.0	2.1	2.3
75–79	0.0	0.0	0.0	0.0
80–84	0.0	0.0	0.0	0.0
85–89	0.0	0.0	0.0	0.0
90–94	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0

Based on scorecard applied to the validation sample.

Figure 10 (USAID "extreme" line): Differences and precision of differences for bootstrapped estimates of households' poverty likelihoods, by sample size, scorecard applied to validation sample

Sample size (n)	Diff.	Difference between estimate and true value		
		Confidence interval (+/− percentage points)		
		90-percent	95-percent	99-percent
2	-2.7	32.9	40.4	56.8
4	-2.5	23.2	29.1	38.7
8	-2.4	17.1	20.8	26.2
16	-2.5	13.1	16.6	21.1
32	-2.3	10.1	12.3	15.2
64	-2.1	7.6	9.1	12.0
128	-1.9	6.2	7.3	9.4
256	-1.7	4.5	5.3	7.5
512	-1.5	3.3	4.0	5.1
1,024	-1.4	2.0	2.4	3.5
2,048	-1.4	1.5	1.7	2.3
4,096	-1.4	1.0	1.2	1.6
8,192	-1.4	0.7	0.9	1.2
16,384	-1.4	0.5	0.6	0.8

Figure 12 (USAID "extreme" line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to validation sample

Sample size (n)		Difference between estimate and true value		
		Confidence interval (+/− percentage points)		
	Diff.	90-percent	95-percent	99-percent
2	-2.7	32.9	40.4	56.8
4	-2.5	23.2	28.5	36.9
8	-2.5	15.7	19.9	25.5
16	-2.4	11.2	13.3	17.1
32	-2.3	8.2	9.6	12.6
64	-2.4	5.7	6.8	9.0
128	-2.5	4.0	4.8	6.1
256	-2.4	2.8	3.3	4.3
512	-2.4	2.1	2.4	3.4
1,024	-2.4	1.3	1.6	2.2
2,048	-2.4	1.0	1.2	1.6
4,096	-2.4	0.7	0.8	1.1
8,192	-2.4	0.5	0.6	0.8
16,384	-2.4	0.3	0.4	0.5

Figure 14 (USAID "extreme" line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: ≥ poverty line mistakenly targeted	Exclusion: ≥ poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.4	10.8	0.0	88.7	89.2	-92.5
5–9	0.9	10.4	0.3	88.4	89.3	-81.4
10–14	1.6	9.6	1.6	87.2	88.8	-57.2
15–19	3.0	8.2	2.3	86.5	89.5	-26.1
20–24	5.0	6.2	6.0	82.8	87.8	42.6
25–29	7.4	3.9	13.6	75.1	82.5	-21.1
30–34	10.1	1.2	20.1	68.6	78.7	-78.7
35–39	10.5	0.8	28.6	60.1	70.6	-154.1
40–44	10.8	0.4	41.9	46.9	57.7	-272.2
45–49	11.0	0.3	50.7	38.1	49.1	-350.1
50–54	11.1	0.1	60.9	27.9	39.0	-441.1
55–59	11.1	0.1	69.8	18.9	30.1	-520.3
60–64	11.1	0.1	75.3	13.5	24.6	-569.0
65–69	11.1	0.1	80.0	8.7	19.8	-611.2
70–74	11.3	0.0	84.3	4.4	15.7	-649.4
75–79	11.3	0.0	86.3	2.4	13.7	-667.1
80–84	11.3	0.0	87.2	1.6	12.8	-674.7
85–89	11.3	0.0	88.0	0.7	12.0	-682.3
90–94	11.3	0.0	88.4	0.3	11.6	-685.8
95–100	11.3	0.0	88.7	0.0	11.3	-688.6

Inclusion, undercoverage, leakage, and exclusion normalized to sum to 100.

Figure 15 (USAID "extreme" line): Households below the poverty line and all households, at a given score or at or below a given score cut-off, scorecard applied to validation sample

Score	Households below poverty line (%)		All households (%)	
	At score	At or below score	At score	At or below score
0–4	100.0	100.0	0.4	0.4
5–9	60.1	74.1	0.8	1.2
10–14	37.3	51.2	2.0	3.2
15–19	66.3	57.2	2.1	5.3
20–24	35.1	45.7	5.7	11.0
25–29	23.5	35.1	10.0	21.0
30–34	29.2	33.3	9.2	30.2
35–39	4.6	26.8	8.9	39.1
40–44	2.7	20.5	13.7	52.7
45–49	1.4	17.8	8.9	61.6
50–54	1.5	15.4	10.4	72.0
55–59	0.0	13.7	8.9	80.9
60–64	0.0	12.9	5.5	86.4
65–69	0.0	12.2	4.7	91.1
70–74	3.1	11.8	4.4	95.6
75–79	0.0	11.5	2.0	97.6
80–84	0.0	11.4	0.9	98.4
85–89	0.0	11.3	0.9	99.3
90–94	0.0	11.3	0.4	99.7
95–100	0.0	11.3	0.3	100.0

$1/Day Poverty Line Tables

Figure 5 ($1/day line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	56.0
5–9	51.8
10–14	22.4
15–19	15.3
20–24	4.0
25–29	5.0
30–34	3.6
35–39	0.0
40–44	1.0
45–49	0.0
50–54	0.0
55–59	0.0
60–64	0.0
65–69	0.0
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Surveyed cases weighted to represent Palestine's households.

Based on the 2005 PECS.

Figure 6 ($1/day line): Derivation of estimated poverty likelihoods associated with scores

Score	Households below poverty line		All households at score		Poverty likelihood (estimated, %)
0–4	236	÷	422	=	56.0
5–9	405	÷	782	=	51.8
10–14	443	÷	1,978	=	22.4
15–19	323	÷	2,105	=	15.3
20–24	226	÷	5,726	=	4.0
25–29	499	÷	9,999	=	5.0
30–34	330	÷	9,156	=	3.6
35–39	0	÷	8,897	=	0.0
40–44	141	÷	13,652	=	1.0
45–49	0	÷	8,888	=	0.0
50–54	0	÷	10,402	=	0.0
55–59	0	÷	8,908	=	0.0
60–64	0	÷	5,481	=	0.0
65–69	0	÷	4,748	=	0.0
70–74	0	÷	4,446	=	0.0
75–79	0	÷	1,987	=	0.0
80–84	0	÷	856	=	0.0
85–89	0	÷	851	=	0.0
90–94	0	÷	392	=	0.0
95–100	0	÷	322	=	0.0

Surveyed cases weighted to represent Palestine's households.

Based on the 2005 PECS.

Figure 8 ($1/day line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample ($n = $ 16,384) from the validation sample, with confidence intervals

| | | Difference between estimate and true value | | |
| | | Confidence interval (+/− percentage points) | | |
Score	Diff.	90-percent	95-percent	99-percent
0–4	11.4	10.2	12.8	16.7
5–9	0.9	7.7	8.8	11.6
10–14	2.7	3.4	4.1	5.6
15–19	-5.8	4.8	5.2	5.7
20–24	-0.6	1.3	1.6	2.0
25–29	-0.1	0.8	1.0	1.4
30–34	-0.8	0.8	1.0	1.4
35–39	-1.1	0.8	0.8	0.9
40–44	1.0	0.0	0.0	0.0
45–49	0.0	0.0	0.0	0.0
50–54	0.0	0.0	0.0	0.0
55–59	0.0	0.0	0.0	0.0
60–64	0.0	0.0	0.0	0.0
65–69	0.0	0.0	0.0	0.0
70–74	0.0	0.0	0.0	0.0
75–79	0.0	0.0	0.0	0.0
80–84	0.0	0.0	0.0	0.0
85–89	0.0	0.0	0.0	0.0
90–94	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0

Based on scorecard applied to the validation sample.

Figure 10 ($1/day line): Differences and precision of differences for bootstrapped estimates of households' poverty likelihoods, by sample size, scorecard applied to validation sample

Sample size (n)	Diff.	Difference between estimate and true value		
		Confidence interval ($+/-$ percentage points)		
		90-percent	95-percent	99-percent
2	-0.1	13.1	26.1	46.3
4	0.0	13.0	16.5	29.9
8	0.0	9.1	12.8	20.6
16	0.1	7.4	10.2	13.5
32	-0.1	6.7	8.1	10.2
64	-0.2	6.2	7.5	10.1
128	-0.2	5.7	6.8	9.2
256	0.0	5.1	6.1	7.7
512	0.2	4.0	4.7	6.2
1,024	0.3	3.1	3.7	4.7
2,048	0.3	2.1	2.5	3.3
4,096	0.4	1.4	1.6	2.2
8,192	0.4	1.0	1.2	1.5
16,384	0.4	0.7	0.9	1.2

Figure 12 ($1/day line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to validation sample

| Sample size (n) | Diff. | Difference between estimate and true value | | |
| | | Confidence interval (+/− percentage points) | | |
		90-percent	95-percent	99-percent
2	-0.1	13.1	26.1	46.3
4	-0.1	12.8	16.2	29.8
8	0.0	8.0	10.6	16.6
16	0.1	5.6	7.6	10.8
32	-0.1	4.1	5.0	7.1
64	-0.1	3.0	3.7	4.8
128	-0.1	2.1	2.6	3.4
256	-0.1	1.5	1.8	2.3
512	-0.1	1.1	1.2	1.8
1,024	-0.1	0.7	0.9	1.1
2,048	-0.1	0.5	0.6	0.9
4,096	-0.1	0.4	0.4	0.6
8,192	-0.1	0.3	0.3	0.4
16,384	-0.1	0.2	0.2	0.3

Figure 14 ($1/day line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: ≥ poverty line mistakenly targeted	Exclusion: ≥ poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.2	2.5	0.2	97.1	97.3	-76.2
5–9	0.6	2.1	0.6	96.7	97.3	-33.2
10–14	1.0	1.7	2.1	95.1	96.2	22.8
15–19	1.5	1.3	3.8	93.4	94.9	-37.5
20–24	1.7	1.1	9.3	87.9	89.6	-237.7
25–29	2.2	0.5	18.8	78.5	80.7	-579.9
30–34	2.6	0.1	27.5	69.7	72.4	-896.4
35–39	2.8	0.0	36.3	60.9	63.7	-1,214.2
40–44	2.8	0.0	50.0	47.3	50.0	-1,708.4
45–49	2.8	0.0	58.8	38.4	41.2	-2,030.2
50–54	2.8	0.0	69.2	28.0	30.8	-2,406.7
55–59	2.8	0.0	78.2	19.1	21.8	-2,729.2
60–64	2.8	0.0	83.6	13.6	16.4	-2,927.6
65–69	2.8	0.0	88.4	8.9	11.6	-3,099.5
70–74	2.8	0.0	92.8	4.4	7.2	-3,260.5
75–79	2.8	0.0	94.8	2.4	5.2	-3,332.4
80–84	2.8	0.0	95.7	1.6	4.3	-3,363.4
85–89	2.8	0.0	96.5	0.7	3.5	-3,394.2
90–94	2.8	0.0	96.9	0.3	3.1	-3,408.4
95–100	2.8	0.0	97.2	0.0	2.8	-3,420.1

Inclusion, undercoverage, leakage, and exclusion normalized to sum to 100.

Figure 15 ($1/day line): Households below the poverty line and all households, at a given score or at or below a given score cut-off, scorecard applied to validation sample

Score	Households below poverty line (%)		All households (%)	
	At score	At or below score	At score	At or below score
0–4	56.0	56.0	0.4	0.4
5–9	51.8	53.3	0.8	1.2
10–14	20.6	33.0	2.0	3.2
15–19	20.8	28.1	2.1	5.3
20–24	3.4	15.3	5.7	11.0
25–29	5.4	10.6	10.0	21.0
30–34	4.5	8.8	9.2	30.2
35–39	1.3	7.1	8.9	39.1
40–44	0.0	5.2	13.7	52.7
45–49	0.0	4.5	8.9	61.6
50–54	0.0	3.8	10.4	72.0
55–59	0.0	3.4	8.9	80.9
60–64	0.0	3.2	5.5	86.4
65–69	0.0	3.0	4.7	91.1
70–74	0.0	2.9	4.4	95.6
75–79	0.0	2.8	2.0	97.6
80–84	0.0	2.8	0.9	98.4
85–89	0.0	2.8	0.9	99.3
90–94	0.0	2.8	0.4	99.7
95–100	0.0	2.8	0.3	100.0

$2/Day Poverty Line Tables

Figure 5 ($2/day line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	100.0
5–9	100.0
10–14	92.4
15–19	69.8
20–24	54.9
25–29	40.7
30–34	37.2
35–39	19.4
40–44	10.7
45–49	6.4
50–54	8.3
55–59	4.4
60–64	4.7
65–69	0.0
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Surveyed cases weighted to represent Palestine's households.

Based on the 2005 PECS.

Figure 6 ($2/day line): Derivation of estimated poverty likelihoods associated with scores

Score	Households below poverty line		All households at score		Poverty likelihood (estimated, %)
0–4	422	÷	422	=	100.0
5–9	782	÷	782	=	100.0
10–14	1,827	÷	1,978	=	92.4
15–19	1,470	÷	2,105	=	69.8
20–24	3,141	÷	5,726	=	54.9
25–29	4,072	÷	9,999	=	40.7
30–34	3,404	÷	9,156	=	37.2
35–39	1,728	÷	8,897	=	19.4
40–44	1,465	÷	13,652	=	10.7
45–49	568	÷	8,888	=	6.4
50–54	865	÷	10,402	=	8.3
55–59	396	÷	8,908	=	4.4
60–64	259	÷	5,481	=	4.7
65–69	0	÷	4,748	=	0.0
70–74	0	÷	4,446	=	0.0
75–79	0	÷	1,987	=	0.0
80–84	0	÷	856	=	0.0
85–89	0	÷	851	=	0.0
90–94	0	÷	392	=	0.0
95–100	0	÷	322	=	0.0

Surveyed cases weighted to represent Palestine's households.

Based on the 2005 PECS.

Figure 8 ($2/day line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample ($n = 16{,}384$) from the validation sample, with confidence intervals

| | | Difference between estimate and true value | | |
| | | Confidence interval (+/− percentage points) | | |
Score	Diff.	90-percent	95-percent	99-percent
0–4	0.0	0.0	0.0	0.0
5–9	0.0	0.0	0.0	0.0
10–14	-2.3	2.1	2.2	2.9
15–19	5.5	4.5	5.3	6.9
20–24	-7.2	4.8	5.1	5.6
25–29	-10.0	6.1	6.2	6.5
30–34	-4.5	3.2	3.5	3.8
35–39	-0.4	1.7	2.1	2.7
40–44	1.1	1.1	1.3	1.7
45–49	-1.1	1.1	1.3	1.8
50–54	3.5	0.8	0.9	1.2
55–59	1.6	0.7	0.8	1.1
60–64	4.7	0.0	0.0	0.0
65–69	0.0	0.0	0.0	0.0
70–74	-2.9	2.0	2.1	2.3
75–79	0.0	0.0	0.0	0.0
80–84	0.0	0.0	0.0	0.0
85–89	0.0	0.0	0.0	0.0
90–94	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0

Based on scorecard applied to the validation sample.

Figure 10 ($2/day line): Differences and precision of differences for bootstrapped estimates of households' poverty likelihoods, by sample size, scorecard applied to validation sample

Sample size (n)	Diff.	Difference between estimate and true value		
		Confidence interval (+/− percentage points)		
		90-percent	95-percent	99-percent
2	-1.8	36.1	46.2	62.9
4	-1.4	27.2	33.8	45.5
8	-1.4	19.8	23.8	31.4
16	-1.2	15.0	18.4	22.1
32	-1.0	10.5	12.7	17.3
64	-0.8	7.1	8.6	11.5
128	-0.8	5.0	6.3	7.9
256	-0.7	3.3	3.9	4.9
512	-0.6	2.1	2.6	3.5
1,024	-0.6	1.4	1.6	2.2
2,048	-0.6	1.0	1.2	1.6
4,096	-0.6	0.7	0.9	1.1
8,192	-0.6	0.5	0.6	0.8
16,384	-0.6	0.3	0.4	0.5

Figure 12 ($2/day line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to validation sample

Sample size (n)	Diff.	Difference between estimate and true value		
		Confidence interval (+/− percentage points)		
		90-percent	95-percent	99-percent
2	-1.8	36.1	46.2	62.9
4	-1.4	27.0	33.2	46.7
8	-1.5	19.0	22.9	29.1
16	-1.3	13.7	15.8	21.5
32	-1.1	9.7	11.6	15.5
64	-1.2	6.6	7.7	10.4
128	-1.3	4.8	5.8	7.3
256	-1.2	3.5	4.2	5.4
512	-1.2	2.5	2.9	4.0
1,024	-1.1	1.7	2.1	2.7
2,048	-1.1	1.2	1.5	1.9
4,096	-1.1	0.8	1.0	1.3
8,192	-1.1	0.6	0.7	0.9
16,384	-1.1	0.4	0.5	0.6

Figure 14 ($2/day line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: ≥ poverty line mistakenly targeted	Exclusion: ≥ poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.4	21.6	0.0	78.0	78.4	-96.2
5–9	1.2	20.8	0.0	78.0	79.2	-89.0
10–14	3.1	18.9	0.1	77.9	80.9	-71.6
15–19	4.5	17.5	0.8	77.2	81.7	-55.4
20–24	8.1	13.9	2.9	75.1	83.3	-12.9
25–29	13.4	8.6	7.6	70.4	83.8	56.4
30–34	17.3	4.7	12.9	65.1	82.4	41.3
35–39	19.0	3.0	20.0	58.0	77.0	8.8
40–44	20.4	1.6	32.4	45.7	66.0	-47.2
45–49	21.0	0.9	40.6	37.4	58.5	-84.5
50–54	21.6	0.4	50.4	27.6	49.2	-129.3
55–59	21.8	0.1	59.1	18.9	40.8	-168.6
60–64	21.8	0.1	64.5	13.5	35.3	-193.5
65–69	21.8	0.1	69.3	8.7	30.6	-215.1
70–74	22.0	0.0	73.6	4.4	26.4	-234.7
75–79	22.0	0.0	75.6	2.4	24.4	-243.8
80–84	22.0	0.0	76.4	1.6	23.6	-247.7
85–89	22.0	0.0	77.3	0.7	22.7	-251.5
90–94	22.0	0.0	77.7	0.3	22.3	-253.3
95–100	22.0	0.0	78.0	0.0	22.0	-254.8

Inclusion, undercoverage, leakage, and exclusion normalized to sum to 100.

Figure 15 ($2/day line): Households below the poverty line and all households, at a given score or at or below a given score cut-off, scorecard applied to validation sample

Score	Households below poverty line (%)		All households (%)	
	At score	At or below score	At score	At or below score
0–4	100.0	100.0	0.4	0.4
5–9	100.0	100.0	0.8	1.2
10–14	93.8	96.2	2.0	3.2
15–19	68.9	85.3	2.1	5.3
20–24	63.2	73.8	5.7	11.0
25–29	52.5	63.7	10.0	21.0
30–34	42.5	57.2	9.2	30.2
35–39	19.7	48.7	8.9	39.1
40–44	9.8	38.6	13.7	52.7
45–49	7.7	34.2	8.9	61.6
50–54	5.2	30.0	10.4	72.0
55–59	3.0	27.0	8.9	80.9
60–64	0.0	25.3	5.5	86.4
65–69	0.0	24.0	4.7	91.1
70–74	3.1	23.0	4.4	95.6
75–79	0.0	22.5	2.0	97.6
80–84	0.0	22.3	0.9	98.4
85–89	0.0	22.1	0.9	99.3
90–94	0.0	22.1	0.4	99.7
95–100	0.0	22.0	0.3	100.0

$3/Day Poverty Line Tables

Figure 5 ($3/day line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	100.0
5–9	100.0
10–14	92.4
15–19	92.2
20–24	92.9
25–29	75.8
30–34	74.2
35–39	55.2
40–44	43.9
45–49	34.6
50–54	19.8
55–59	10.6
60–64	14.4
65–69	7.1
70–74	4.3
75–79	5.8
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Surveyed cases weighted to represent Palestine's households.

Based on the 2005 PECS.

Figure 6 ($3/day line): Derivation of estimated poverty likelihoods associated with scores

Score	Households below poverty line		All households at score		Poverty likelihood (estimated, %)
0–4	422	÷	422	=	100.0
5–9	782	÷	782	=	100.0
10–14	1,827	÷	1,978	=	92.4
15–19	1,942	÷	2,105	=	92.2
20–24	5,318	÷	5,726	=	92.9
25–29	7,577	÷	9,999	=	75.8
30–34	6,794	÷	9,156	=	74.2
35–39	4,914	÷	8,897	=	55.2
40–44	5,986	÷	13,652	=	43.9
45–49	3,073	÷	8,888	=	34.6
50–54	2,064	÷	10,402	=	19.8
55–59	946	÷	8,908	=	10.6
60–64	790	÷	5,481	=	14.4
65–69	339	÷	4,748	=	7.1
70–74	193	÷	4,446	=	4.3
75–79	116	÷	1,987	=	5.8
80–84	0	÷	856	=	0.0
85–89	0	÷	851	=	0.0
90–94	0	÷	392	=	0.0
95–100	0	÷	322	=	0.0

Surveyed cases weighted to represent Palestine's households.

Based on the 2005 PECS.

Figure 8 ($3/day line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample ($n = 16{,}384$) from the validation sample, with confidence intervals

| | | Difference between estimate and true value | | |
| | | Confidence interval ($+/-$ percentage points) | | |
Score	Diff.	90-percent	95-percent	99-percent
0–4	0.0	0.0	0.0	0.0
5–9	0.0	0.0	0.0	0.0
10–14	-7.6	3.8	3.8	3.8
15–19	-7.8	3.9	3.9	3.9
20–24	8.6	2.0	2.3	2.9
25–29	-1.2	1.8	2.0	2.8
30–34	0.5	2.0	2.3	3.0
35–39	-6.4	4.3	4.6	5.0
40–44	-0.7	1.8	2.1	2.8
45–49	2.5	2.0	2.4	3.0
50–54	1.7	1.6	1.9	2.6
55–59	-7.5	4.6	4.7	5.1
60–64	4.4	1.6	1.9	2.6
65–69	4.5	0.9	1.1	1.5
70–74	-5.1	3.4	3.7	4.1
75–79	-3.6	3.3	3.5	4.3
80–84	0.0	0.0	0.0	0.0
85–89	0.0	0.0	0.0	0.0
90–94	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0

Based on scorecard applied to the validation sample.

Figure 10 ($3/day line): Differences and precision of differences for bootstrapped estimates of households' poverty likelihoods, by sample size, scorecard applied to validation sample

Sample size (n)	Diff.	Difference between estimate and true value Confidence interval (+/− percentage points)		
		90-percent	95-percent	99-percent
2	-1.6	46.0	57.1	70.4
4	-1.1	34.4	40.2	50.4
8	-0.9	24.4	29.1	39.5
16	-0.9	17.5	20.6	27.9
32	-0.5	11.7	13.7	17.6
64	-0.8	7.4	9.1	12.0
128	-1.0	4.8	5.6	7.7
256	-1.0	3.0	3.6	4.9
512	-0.9	2.0	2.4	3.2
1,024	-0.9	1.4	1.6	2.1
2,048	-0.9	1.0	1.2	1.5
4,096	-0.9	0.6	0.8	0.9
8,192	-0.9	0.5	0.6	0.7
16,384	-0.9	0.3	0.4	0.5

Figure 12 ($3/day line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to validation sample

Sample size (n)	Diff.	Difference between estimate and true value		
		Confidence interval (+/− percentage points)		
		90-percent	95-percent	99-percent
2	-1.6	46.0	57.1	70.4
4	-1.0	33.8	39.4	50.2
8	-0.8	23.6	29.3	38.0
16	-0.9	17.6	20.6	26.2
32	-0.5	11.9	14.0	18.6
64	-0.6	8.4	9.9	12.9
128	-0.7	5.9	6.9	9.0
256	-0.7	4.1	4.9	6.4
512	-0.7	2.9	3.6	4.7
1,024	-0.7	2.1	2.5	3.2
2,048	-0.6	1.5	1.8	2.3
4,096	-0.6	1.0	1.2	1.6
8,192	-0.6	0.8	0.9	1.3
16,384	-0.7	0.5	0.6	0.8

Figure 14 ($3/day line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: ≥ poverty line mistakenly targeted	Exclusion: ≥ poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.4	43.9	0.0	55.7	56.1	-98.1
5–9	1.2	43.1	0.0	55.7	56.9	-94.6
10–14	3.2	41.1	0.0	55.7	58.9	-85.6
15–19	5.3	39.0	0.0	55.7	61.0	-76.1
20–24	10.1	34.2	0.9	54.8	64.9	-52.3
25–29	18.0	26.3	3.0	52.7	70.6	-12.0
30–34	24.8	19.5	5.4	50.3	75.2	24.1
35–39	30.4	13.8	8.6	47.1	77.5	56.9
40–44	36.6	7.7	16.1	39.6	76.2	63.6
45–49	39.5	4.8	22.1	33.6	73.1	50.1
50–54	41.3	3.0	30.7	25.0	66.3	30.7
55–59	43.0	1.3	37.9	17.8	60.8	14.4
60–64	43.6	0.7	42.8	12.9	56.4	3.3
65–69	43.7	0.6	47.4	8.3	51.9	-7.1
70–74	44.1	0.2	51.5	4.2	48.4	-16.2
75–79	44.3	0.0	53.3	2.4	46.7	-20.3
80–84	44.3	0.0	54.1	1.6	45.9	-22.2
85–89	44.3	0.0	55.0	0.7	45.0	-24.1
90–94	44.3	0.0	55.4	0.3	44.6	-25.0
95–100	44.3	0.0	55.7	0.0	44.3	-25.8

Inclusion, undercoverage, leakage, and exclusion normalized to sum to 100.

Figure 15 ($3/day line): Households below the poverty line and all households, at a given score or at or below a given score cut-off, scorecard applied to validation sample

| Score | Households below poverty line (%) | | All households (%) | |
	At score	At or below score	At score	At or below score
0–4	100.0	100.0	0.4	0.4
5–9	100.0	100.0	0.8	1.2
10–14	100.0	100.0	2.0	3.2
15–19	100.0	100.0	2.1	5.3
20–24	84.5	91.9	5.7	11.0
25–29	78.5	85.6	10.0	21.0
30–34	74.6	82.2	9.2	30.2
35–39	63.4	77.9	8.9	39.1
40–44	45.1	69.4	13.7	52.7
45–49	32.4	64.1	8.9	61.6
50–54	17.4	57.3	10.4	72.0
55–59	19.0	53.1	8.9	80.9
60–64	10.3	50.4	5.5	86.4
65–69	2.9	47.9	4.7	91.1
70–74	9.7	46.2	4.4	95.6
75–79	8.5	45.4	2.0	97.6
80–84	0.0	45.0	0.9	98.4
85–89	0.0	44.6	0.9	99.3
90–94	0.0	44.4	0.4	99.7
95–100	0.0	44.3	0.3	100.0