Poverty Scorecards: Lessons from a Microlender in Bosnia-Herzegovina

February 19, 2006

Mark Schreiner
Center for Social Development, Washington University in Saint Louis
Campus Box 1196, One Brookings Drive, Saint Louis, MO 63130-4130, U.S.A.
Phones: (314) 481-9788 or (314) 752-9022
schreiner@gwbmail.wustl.edu

Michal Matul and Ewa Pawlak
Microfinance Centre for Central and Eastern Europe and the New Independent States
ul. Koszykowa 60/62 m. 52, 00-673 Warsaw, Poland
michal@mfc.org.pl, ewa.pawlak@indigo.pl

Sean Kline
Freedom from Hunger
1644 DaVinci Court, Davis, CA 95616, U.S.A
skline@freedomfromhunger.org

Key words: Eastern Europe, Bosnia-Herzegovina, poverty measurement, microfinance

Abstract

How poor are participants in development projects? This paper analyzes how well a simple scorecard identifies poor clients at a microlender in Bosnia-Herzegovina. The scorecard effectively ranks clients by relative poverty and also identifies the likelihood that a client is poor by an absolute standard. The score tracks poverty more closely than loan size, microfinance's traditional poverty indicator. Overall, poverty scorecards are a simple, inexpensive way for microlenders—or any other development entity—to target the poor, track changes in poverty over time, manage poverty outreach, and report on clients' absolute poverty.

Acknowledgements

We are grateful for funding from the Consultative Group to Assist the Poorest and from the Ford Foundation. The poverty scorecard was based on data from the 2001 Living Standards Measurement Survey by the World Bank, the State Agency for Statistics in Bosnia-Herzegovina, the Federation Statistical Institute, and the Republika Srpska Statistical Institute.

Poverty Scorecards: Lessons from a Microlender in Bosnia-Herzegovina

Abstract

How poor are participants in development projects? This paper analyzes how well a simple scorecard identifies poor clients at a microlender in Bosnia-Herzegovina (BiH). The scorecard effectively ranks clients by relative poverty and also identifies the likelihood that a client is poor by an absolute standard. The score tracks poverty more closely than loan size, microfinance's traditional poverty indicator. Overall, poverty scorecards are a simple, inexpensive way for microlenders—or any other development entity—to target the poor, track changes in poverty over time, manage poverty outreach, and report on clients' absolute poverty.

Poverty Scoring:
Lessons from a Microlender in Bosnia-Herzegovina

1. INTRODUCTION

Microfinance in particular—and development assistance in general—aims to help the poor. But what share of participants are poor? Expenditure surveys measure poverty directly, but they are expensive. A less-expensive approach estimates expenditure-based poverty with a scorecard using simple-to-collect, non-expenditure indicators (Zeller, 2004; Hatch and Frederick, 1998). While these scores are correlated with poverty, no one knows how well they are correlated.[1] This paper examines the power of a poverty scorecard at Prizma, a microlender in Bosnia-Herzegovina (BiH). The scorecard estimates the likelihood that a given client is poor, and averaging each client's poverty likelihood gives an estimate of the poverty rate among the microlender's clients.

The poverty scorecard does a good job of identifying poor clients. Because the scorecard was derived from a Living Standard Measurement Survey (World Bank *et al.*, 2002), it can also produce estimates of absolute poverty rates. About 17.9 percent of Prizma's new borrowers were poor (versus 19.3 percent for all BiH). Compared with loan size—the most common poverty proxy in microfinance, the scorecard identifies poor clients much better.

These results have two caveats. First, some scorecard indicators were not collected exactly as in the national LSMS survey, so the estimates are biased to some unknown degree. Second, the scorecard does not completely control for the fact that Prizma's clients are not a random sample of the population of BiH.

Prizma's poverty scorecard is both powerful and practical, with inexpensive-to-collect indicators and weights that allow field workers to compute scores on paper. The example here is in microfinance, but poverty scorecards could be applied just as well in other development contexts.

This paper describes Prizma's poverty scorecard and its construction, measures the scorecard's power, and compares it with a scorecard based on loan size. The conclusion discusses lessons for poverty scorecards in microfinance and other contexts.

II. CONSTRUCTING A POVERTY SCORECARD

Constructing Prizma's poverty scorecard involved:

- Measuring the absolute, expenditure-based poverty status of households in a national random sample
- Selecting non-expenditure indicators that were both inexpensive-to-collect and correlated with absolute, expenditure-based poverty status
- Constructing a scorecard by assigning weights to the non-expenditure indicators to reflect their correlation with expenditure-based poverty status
- Adding up the weighted non-expenditure indicators to produce poverty scores for the surveyed households

- Collecting from clients the non-expenditure indicators used in the scorecard and using them to compute the clients' poverty scores
- Defining the poverty likelihood of a client with a given poverty score as the observed poverty rate among surveyed households with the same score
- Defining the overall poverty rate as clients' average poverty likelihood
- Checking that poverty scores made sense for different branches, products, and geographic areas

This process rests on two basic assumptions. The first is that Prizma's clients—like the surveyed households—are a random sample from the population of BiH. The second is that the relationship between non-expediture indicators and expenditure-based poverty status does not change through time. Consequences of violating these assumptions are discussed later.

(a) An expenditure-based measure of absolute poverty

Poverty status was derived from BiH's 2001 Living Standards Measurement Survey. The LSMS recorded expenditure and a wide range of other data for a national random sample. A household was *poor* if annual per capita consumption (adjusted for the local cost of living) was less than 2,200 Convertible Marks (World Bank *et al.*, 2002). At purchasing-power parity, this poverty line was about $14 per person per day.[2] The overall poverty rate in BiH was 19.3 percent.

(b) Indicator selection

Scorecard builders (Matul and Kline, 2003) selected indicators that:

- Correlated strongly with poverty status, both in the past and future
- Appeared in the national survey, enabling linkage with an absolute poverty line
- Kept data-collection costs low:
    - Already collected as part of the loan evaluation, or easy to start to collect
    - Did not make clients or loan officers uncomfortable
- Elicited truthful reports that an internal auditor could verify
- Took different values across clients
- Took different values for a given client as poverty changes over time

Analysts brainstormed a long list of candidate indicators, drawing on their country knowledge and poverty studies (Dunn and Tvrtkovic, 2003; Prism Research, 2003; Bisogno and Chong, 2002; World Bank *et al.*, 2002). The list was narrowed with the criteria above and input from managers, staff, and client focus groups.

(i) Indicators directly linked with the national survey

Some scorecard indicators had direct analogs in the LSMS. Car ownership, for example, was strongly correlated with expenditure-based poverty: 11 percent of car owners were poor, versus 26 percent of non-owners (Figure 1). Car ownership also varied across households (55 percent were owners, 45 percent non-owners). Prizma found that clients were comfortable reporting whether they owned a car, and the indicator promised to be useful for tracking changes in poverty over time.

Following this same process, scorecard builders selected indicators for:

- Education of the female household head/spouse/partner. In the national survey, female education was highly correlated with overall household education. Also, until recently all Prizma clients are women, so asking only about female's education simplified data collection. Among the 64 percent of surveyed households whose female head had only a primary education, 24 percent were poor. Among the other 36 percent of households, 11 percent were poor (Figure 1)
- Household size. Poverty was strongly correlated with household size, with larger households being poorer
- Stereo CD ownership. 78 percent of households did not own a stereo CD player, and 23 percent of them were poor. In the other 22 percent, 8 percent were poor

Prizma collected these four indicators exactly as in the national survey. Thus, a scorecard using only these indicators can be directly benchmarked to the LSMS's absolute, expenditure-based measure of poverty status.

(ii) Indicators not directly linked with the national survey

The scorecard uses three other indicators—location of residence, frequency of eating meat, and frequency of eating sweets—that Prizma collected differently than the LSMS. In strict terms, scores based on these indicators cannot be linked to the survey's poverty measure. Still, these indicators were highly correlated with poverty, so they still help rank clients by relative poverty.

For location of residence, Prizma classified clients as urban or rural by population. The national survey, however, assigned location status by municipality, even though many municipalities have both rural and urban areas. In the survey, about 21 percent of people in rural municipalities were poor versus 13 percent in urban municipalities (Figure 1). This need not, however, imply anything about the poverty of clients whose location of residence is defined by population. The estimates of overall poverty rates in this paper assume that Prizma's definition of rural/urban is equivalent to the LSMS definition.

A poverty-assessment survey (Henry *et al.*, 2003) for Prizma found that the frequency of eating meats and sweets was highly correlated with poverty (Prism Research, 2003). The LSMS, however, recorded food spending, not frequency. Prizma found it more practical to ask about frequency: the times per week the household eats meat and the times per week the household eats sweets (usually cakes) with the main meal. If all households were the same size and if all people ate the same amount, then frequency (measured by Prizma) would be perfectly correlated with spending (measured by the survey). In fact, larger households can spend more on meat (or sweets) even if they eat less frequently, and different people eat different amounts. Thus, Prizma's indicator is not equivalent to the survey indicator, breaking the direct link between the scorecard and the expenditure-based poverty benchmark.

Knowing this, the scorecard builders divided the survey distribution of meat spending into three classes so that its distribution matched that of a sample of Prizma's clients in terms of “rarely” (0–2 times per week), “sometimes” (3–5), and “often” (6–7). Figure 1 shows that meat spending was highly correlated with poverty in the survey: the 42 percent of the lowest spenders were poor, versus 19 percent for those in the middle and 4 percent for the highest spenders. (Spending on sweets was also correlated with poverty, but less strongly.) Still, the correlation between spending (measured in the survey) and frequency (measured by Prizma) is unknown. The overall poverty rates in this paper assume that the correlation is perfect.

(iii) Excluded indicators

The scorecard does not include all indicators that Prizma collects and that are in the survey. For example, an indicator for single mothers was left out because female- and male-headed households with children had about the same poverty rate.

Scorecard builders also considered—but ultimately rejected—some survey indicators that were strongly correlated with poverty but that fell short on other criteria. For example, refugee status in 2001 was strongly correlated with poverty, but Prizma’s managers believe that this correlation is disappearing. Thus, including the indicator would cause the scorecard to overestimate the poverty likelihood of refugees.

Likewise, the survey found that the unemployed were more likely to be poor. The survey’s aggregate unemployment figure, however, was impossibly high, probably because many part-time or unregistered workers were counted as unemployed. Because Prizma would (correctly) record such workers as employed, a scorecard that included employment would underestimate poverty likelihood.

Finally, while television ownership was highly correlated with poverty, about 96 percent of Prizma's clients owned TVs. With so little variation across clients (and less variation through time), TV ownership does not help rank clients by poverty.

(iii) Lessons for the selection of poverty indicators

Selecting poverty indicators is not just a statistical exercise. Even indicators that are strongly correlated with poverty in a national survey might produce misleading scores if they do not really measure what they say they measure, if the relationship between indicators and poverty is changing, or if the lender records the indicators differently than the survey. Other indicators are too costly to collect accurately. All in all, poverty scorecards depend more on data quality than statistical sophistication.

Building poverty scorecards requires "domain expertise", that is, knowledge of the local context and of the specific development project. Feedback from front-line staff is also key, as are pilot tests. The power of the poverty scorecard examined here comes less from the specific weights assigned to the indicators than from someone's realizing that the standard in BiH is to eat cake with the main meal and that the culture's love of music makes the lack of a stereo CD player an indicator of poverty.

(c) Indicator weights

Four of the seven poverty indicators had "Yes/No" answers. A client either did or did not own a car, have more than a primary education, own a stereo CD, or live in an urban area. Weights of zero (0) were given to values correlated with greater poverty in the survey, and weights of one (1) were given to values correlated with less poverty.

The two food indicators (frequency of eating meat and sweets) had values of “rarely”, “sometimes”, or “often” with weights of 0, 1, or 2, again reflecting survey correlations. Finally, household size was divided into six classes (1, 2, 3, 4, 5, and 6 or more).

Figure 2 lists the scorecard’s indicators, values, and weights. (The four-indicator scorecard will be discussed later.) Weights were derived from a logit regression and then adjusted so that:

- All weights are integers
- All weights are positive
- Scores range from 0 (most likely poor) to 100 (least likely poor).

The small number of indicators, their simple form, and this weighting scheme allows field workers to compute scores on paper. But can such a simple scorecard accurately identify poor participants? Both theory and practice support such simple scorecards. They have been used by banks to predict creditworthiness, hospitals to identify at-risk pregnancies, phone companies to predict late bill-payment, and colleges to screen potential matriculants (Lovie and Lovie, 1986; Kolesar and Showers, 1985; Stillwell, Barron, and Edwards, 1983; Dawes, 1979; Myers and Forgy, 1963). Wainer (1976) shows mathematically why simple scorecards can work. Simple scorecards are also robust to data problems and—perhaps most important—help users see where the scores come from.

III. POWER TO IDENTIFY THE POOR

This section tests scorecard power by checking whether it tends to assign lower scores to households in the LSMS who were poor.

Given a score, Figure 3 shows the number of households with that score, the share of household with that score, and the percentage of those households who were poor in the national survey. For example, 46.2 households (households were weighted for national representativeness) had scores from 0 to 3, and 44 were poor. Thus, scores of 0 to 3 were associated with a poverty likelihood of 95.3 percent (44) 46.2). Households with scores of 0 to 3 represented 0.9 percent of all households but 4.5 percent of poor households.

Among households with scores from 4 to 15, 80.2 percent (229.2 of 285.8) were poor. These represented 5.7 percent of all households and 23.5 percent of the poor.

The scorecard also assigned higher scores (and lower poverty likelihoods) to non-poor households. For example, 1,837.4 households had scores from 45 to 100, and 22.5 of them (1.2 percent) were poor. Overall, the scorecard assigned lower scores (and higher poverty likelihoods) to poor households.

A formal measure of scorecard accuracy is the “Power Curve” (Figure 4). The line arcing toward the northwest corner shows power to identify the poor. It traces out the share of poor households with a given score or less (vertical axis) against the share of all households with that score or less (horizontal axis). For example, 26.6 percent of poor households had a score of 1 or less, while 6.9 percent of all households (poor and non-poor) had a score of 1 or less. The greater the power to identify the poor, the more the curve approaches the top and left borders. For example, the curve almost touches the top border for the 20 percent of households with the highest scores, as the scorecard correctly identifies almost all of these households as non-poor.

In Figure 4, the line arcing toward the southeast corner shows the power to identify non-poor households. The closer this curve is to the bottom and right borders, the greater the power.

The greater the area between the two curves, the greater the power. One measure of this is the Kolmogorov-Smirnov distance, the vertical distance between the poor and non-poor curves.[3] For Prizma's poverty scorecard, the maximum KS is 0.60.[4] Given its simplicity, the scorecard overall is remarkably powerful.

How important is each indicator? Figure 5 measures importance (normalized on a scale of 0 to 100) as the reduction in the log-likelihood in the logit due to removing a given indicator while keeping all others (Brieman, 2001). Household size is the most powerful indicator, followed by frequency of eating meat, frequency of eating sweets, and car ownership. In BiH, a two-indicator scorecard with meat consumption and household size would likely be both simple and powerful. Three indicators—ownership of a stereo CD player, location of residence, and education—contribute little.

Except for household size, the most-important indicators may change if poverty changes. This suggests that the scorecard can track changes in poverty over time.

Prizma's poverty scorecard is simple, inexpensive, and powerful. It effectively assigns lower scores to clients who are more likely to be poor and higher scores to clients who are less likely to be poor.

IV. OVERALL POVERTY RATE

The poverty scorecard ranks clients by relative poverty. These ranks can help managers improve targeting, track changes in poverty status over time, and manage poverty outreach. Ranking clients requires that lower scores be associated with higher poverty likelihoods, but it does not require knowing the exact likelihoods.

Donors, however, want measures of absolute poverty, and this does require exact likelihoods. If a poverty scorecard is based on an expenditure survey, these likelihoods are known (at least to the extent that indicators can be directly linked to the survey). A given client's poverty likelihood is the share of clients in the national survey with that score who were poor. Overall poverty rates are then clients' average poverty likelihoods.

(a) Why measure rates of absolute poverty (and how)

Poverty ranks are sufficient for managers but not for donors. When alloting funds across organizations, absolute measures are required to compare apples with apples. Absolute measures also act as a reality check on claims of poverty outreach by microfinance advocates. Furthermore, they create incentives for managers to innovate to reach more and poorer clients (Dunford, 2002a). For example, publishing comparable (that is, absolute) measures of poverty for peer groups of microlenders would increase pressure to improve poverty outreach.

In addition, all recipients of microenterprise assistance from the U.S. Agency for International Development must—as of October 2005—report the share of clients who are "very poor", defined as living on less than a dollar per day or being among the poorest half of people below the country-specific poverty line. The U.S. Congress (Public Law 108-31) requires that these measures be objective (linked with an expenditure-based poverty line), quantitative (not "more or less poor" but "above or below the poverty line") and low-cost (Zeller, 2004).

There are three broad approaches to meeting these goals. The first—pioneered by Freedom from Hunger—uses Lot Quality Assurance Sampling (Davis, 2002; MkNelly *et al.*, 2002). It uses an expenditure survey with a small sample of clients to estimate the probability that at least 50 percent of all clients are poor. Lot Quality Assurance Sampling has high per-client costs because of the survey but low total costs because very few clients are surveyed.

The second approach is that taken by Prizma described in this paper. It produces objective, quantitative poverty measures without additional surveys. It also scores all clients (not just a sample), so it can be used for targeting. This approach assumes that clients are selected at random. If they are not, then it assumes that the scorecard indicators control for non-random differences between clients and non-clients that affect both poverty status and the probability of being selected as a client. This assumption about "selection effects" will be revisited below.

The third approach—used by the IRIS Center at the University of Maryland (Zeller, 2004) to help microfinance organizations meet Congress's mandate—is like Prizma's approach except the scorecard is based not on an existing expenditure survey but rather a new special-purpose expenditure survey on a national random sample. Doing a new survey allows the inclusion of non-expenditure indicators that do not appear in existing surveys. Of course, doing a new survey is also costly.

Only Lot Quality Assurance Sampling measures client poverty directly and so avoids bias due to "selection effects". These effects occur because clients are self-selected (they choose to apply to programs) and program-selected (programs choose which applicants to accept). Both types of selection are partly based on client characteristics (for example, "work ethic", "good looks", or "business sector") correlated with poverty but omitted from the scorecard. Thus, a client and a non-client can have the same score—and even the same values for all indicators—but different poverty statuses.

If an expenditure survey included both clients and non-clients, it could measure selection bias as the difference in poverty likelihood between clients and non-clients with the same poverty scores. Poverty rates could then be adjusted for "selection effects".

Another alternative is to ask in the expenditure survey about the presence of formal loans. This indicator could then be related to poverty and included in the scorecard. All microloan clients would have formal loans, and this increases clients' scores (assuming omitted indicators are positively correlated with selection as a client and negatively correlated with poverty), building-in an adjustment for selection bias.

Overall, the three approaches reflect trade-offs between different goals. Lot Quality Assurance Sampling checks whether a given standard of poverty outreach is met, but it is probably less accurate for estimating an overall poverty rate, and it cannot track changes in poverty status for a large number of clients over time. The approaches of Prizma and IRIS fulfill all three goals, and they can also help to target services. Compared with Lot Quality Assurance Sampling, however, they may be more costly. Overall, Prizma and IRIS are quite similar and have similar on-going costs, but IRIS has greater up-front costs (because its conducts a survey) and offers greater accuracy (because it provides indicators absent from existing surveys).

(b) Overall poverty rate

Prizma's overall poverty rate is the average poverty likelihood of its clients. Loan officers collected scorecard indicators for 5,177 first-time borrowers from December 2003 to September 2004. The poverty likelihood of each client is defined as the poverty likelihood of households in the national survey with the same score as the client. For example, 65.0 percent of surveyed households with a score of 16 were poor, so Prizma clients with a score of 16 had a poverty likelihood of 65.0 percent.

Figure 6 shows the distribution of clients by score. The poverty rate for Prizma was 17.9 percent. Given that the national figure is 19.3 percent, is this poverty outreach high or low? There is no simple answer, and the national average may not be an appropriate benchmark. After all, the distribution of demand by creditworthy borrowers who are not served by other formal lenders probably is not be uniform over the distribution of poverty. Also, Prizma's poverty outreach may be high compared with the (unknown) poverty outreach other microlenders[5] in BiH or compared with the (unknown) poverty outreach that is sustainable. In any case, Prizma has an explicit mission to serve the poor, and measuring poverty outreach helps managers look for new ways to improve.

(c) Disaggregated poverty rates

While external stakeholders focus on the overall poverty rate, managers also want to look at poverty rates by loan product and by branch because this information might suggest ways to deepen outreach.

The poverty rate at Prizma's Sarajevo branch is 27.1 percent, four times the 6.1-percent rate for Banja Luka (Figure 7). The Zenica branch had a rate of 23.9 percent, versus 13.0 and 14.5 percent for the Mostar and Bihać branches. The reasons for these differences are not immediately obvious. The Republic of Srpska is generally poorer than the the Federation of Bosnia and Herzegovina, but the only branch entirely in the Republic of Srpska—Banja Luka—has the smallest concentration of poor clients. The branches in Sarajevo, Mostar, and Tuzla are all in the the Federation of Bosnia and Herzegovina but also serve some Srpska regions. The Sarajevo branch—with the highest density of poor clients—faces the most competition and thus may have deeper poverty outreach because less-poor borrowers are already served by competitors. The Sarajevo branch also serves some low-income suburbs in the Republic of Srpska.

Figure 8 breaks down poverty rates by loan product. More than 90 percent of new borrowers are "enterprise" borrowers who received group loans or "basic needs" borrowers who receive individual, small, short, unrestricted loans based on the guarantee of a household member with a salaried job. Basic-needs loans are often used for emergencies, and basic-needs borrowers were more likely to be poor (20.4 percent) than enterprise borrowers (15.9 percent). This might be due to the group-individual distinction or the enterprise/emergency distinction. Either way, managers might now investigate the reasons and perhaps take advantage of them to improve outreach.

Overall, poverty outreach varies more by branch than by product. Also, newer/smaller/non-growing branches (those with fewer new clients) had lower concentrations of poverty, perhaps because older/larger/growing branches face more pressure (or are more able, due to experience) to go beyond less-poor clients.

(d) Poverty rates with a fully benchmarked scorecard

The poverty estimates above assume that all the indicators collected by Prizma can be linked directly to the national survey. As discussed earlier, however, this is not the case for location of residence and the frequency of consumption of meats and sweets. How well does the scorecard work if restricted to only the four benchmarked indicators (ownership of cars and stereo CDs, education, and household size)?

Weights for the fully benchmarked scorecard are in Figure 2. The power curve in Figure 9 shows that the fully benchmarked scorecard identifies those most-likely and least-likely to be poor almost as well the seven-indicator scorecard. The fully benchmarked scorecard is, however, less accurate for "middle" scores (assuming—

perhaps incorrectly—that the seven-indicator scorecard is accurate). For the overall poverty rate, the four-indicator scorecard gives 19.5 percent (Figure 10).

V. LOAN SIZE AS A PROXY FOR POVERTY

Does loan size correlated with poverty? The amount disbursed is the most common poverty indicator used in microfinance, although its accuracy is unknown (Dunford, 2002b).

For Prizma clients, there is no direct, expenditure-based measure of poverty status, only the poverty score. Thus, the available data do not provide a direct way to link poverty with loan size. It is possible to test, however, the strength of the link between loan size and poverty likelihood as derived from Prizma's scorecard.

Figure 11 shows a loan-size-only scorecard and the poverty likelihoods for each score, based on the seven-indicator scorecard for Prizma's clients. For example, 23.8 percent of clients with loans from 0 to 400 Convertible Marks were poor. The overall poverty likelihood is 17.9 percent, by definition equal to the seven-indicator scorecard.

Estimates from the loan-size-only scorecard and the seven-indicator scorecard are not highly correlated (Figure 12). For example, 40 percent of all poor clients had scores in the lowest decile in the seven-indicator scorecard, while the lowest decile had 12 percent of all poor clients for the loan-size-only scorecard. On the other end of the distribution, the highest quartile had 1 percent of all poor clients with seven indicators but 20 percent of poor clients for loan-size-only. The loan-size-only scorecard is not highly correlated with the seven-indicator scorecard.

VI. LESSONS FOR POVERTY MEASUREMENT

The poverty scorecard examined here features simple weights and seven inexpensive-to-collect indicators. It effectively identifies poor clients without incurring the cost of directly measuring expenditure. By ranking clients by relative poverty, it helps managers target the poor, track changes in poverty, and manage poverty outreach. By relating scores to poverty status as measured in an expenditure survey, the scorecard also informs donors about clients' absolute poverty. As proxy for poverty, the scorecard works much better than loan size. The scorecard is not specific to Prizma and so could be used by other microlenders in BiH (or indeed for any other poverty-measurement purpose in that country). This concluding section presents nine broad lessons from this analysis for microfinance and development in general.

First, poverty scorecards can work, and they need not be complex or costly. Ranking clients by relative poverty requires finding yes/no (or low/average/high) indicators correlated with poverty. Most microfinance organizations (and many other development projects) already collect several such indicators, and—if desired—they might be able to collect a few more without overburdening field workers and clients.

Second, if a scorecard is derived from an expenditure survey, then it can estimate poverty rates based on absolute benchmarks. Because clients are self-selected and program-selected, however, such estimates are biased (usually upwards). Reducing bias requires including many indicators and/or surveying both clients and non-clients.

Third, both theory and experience provide support for simple weighting schemes. In general, data quantity and data quality matter more than statistical sophistication. After all, no amount of manipulation can substitute for a missing indicator or squeeze meaning from carelessly recorded data. Collecting good data and monitoring its quality is difficult, but the long-term reward will only increase as scorecards—for poverty, repayment behavior, drop-out, and other uncertain future outcomes—spread.

Fourth, programs might use two scorecards, the first with more indicators (not all fully benchmarked) that managers can use to rank individual clients by relative poverty, and a second with fewer indicators (all fully benchmarked) that donors can use to estimate absolute poverty rates for all clients. Including non-benchmarked indicators makes the larger scorecard more accurate and so more useful targeting and tracking.

Fifth, because selection biases are stronger with fewer indicators, smaller scorecards will tend to overstate poverty rates. Unfortunately, it is tempting to use small scorecards, not only because they provide higher estimates of poverty rates but also because they are simpler and less costly. Indeed, unless an organization plans to use poverty scorecards for management purposes, it will have weak incentives to collect quality data and build an accurate scorecard. If donors want accurate poverty rates, they should support poverty scorecards that are useful for management purposes.

Sixth, loan size is correlated with poverty likelihood, but—at least in the case of Prizma—not very strongly.

Seventh, “domain knowledge” (of the specific country, intervention, and program) is key. For example, if almost everyone owns a home, then home ownership is not a useful poverty indicator. Likewise, religion or ethnicity might be highly correlated with poverty but difficult to record without undermining client rapport. Even within a given organization, a single scorecard (if it has few indicators) might not work for all regions or services, perhaps indicating customized scorecards for different client segments.

Eight, there is nothing about poverty scorecards that is specific or unique to microfinance. Indeed, a scorecard that includes enough relevant indicators might serve all the poverty-measurement purposes of development projects in a given country. Indeed, the benefit-cost ratio would be very large if the World Bank and national statistical agencies (when they do LSMS or other expenditure surveys) would assign a few person-weeks to building a poverty scorecard based on their expenditure data.

Ninth and most important, poverty scoring can promote a organizational culture of intentional, explicit management of poverty outreach. Equipped with poverty scores, managers no longer must guess how poor their clients are nor how their poverty status is changing over time. Instead, they can use a standard yardstick to reward branches and field workers who improve proverty outreach. Lack of evidence about poverty outreach no longer supports a business-as-usual complacency, and greater knowledge may spur innovation. Managers cannot hide behind ignorance when they report subjective (and perhaps sanguine, see Dunford, 2002a) estimates of overall poverty rates. Measurement feeds management, and boards and managers equipped with poverty scores cannot help but increase their consideration of poverty outreach.

REFERENCES

Bisogno, M., & Chong A. (2002). Poverty and Inequality in Bosnia and Herzegovina after the Civil War, World Development, 30, 61–76.

Brieman, L. (2001). "Statistical Modeling: Two Cultures", Statistical Science, 16(3), 199–231.

Davis, R. (2002). Lot Quality Assurance Sampling (LQAS) for Microfinance Institutions: A Management Tool to Efficiently Assess Poverty Outreach, Davis, CA: Freedom from Hunger.

Dawes, R. M. (1979). The Robust Beauty of Improper Linear Models in Decision Making, American Psychologist, 34(7), 571–582.

Dunford, C. (2002a). Why Set a Threshold for Service Orientation to the Very Poor? Davis, CA: Freedom from Hunger, `http://www.ffhtechnical.org/publications/pdfs/tresholdsapr02.pdf`.

Dunford, C. (2002b). What's Wrong with Loan Size?, Davis, CA: Freedom from Hunger, `http://www.ffhtechnical.org/publications/pdfs/loansize0302.pdf`.

Dunn, E., & Tvrtkovic, J. (2003). Microfinance Clients in Bosnia and Herzegovina: Report on Baseline Survey, Foundation for Sustainable Development of the Federation of Bosnia and Herzegovina, `http://www.odraz.ba/Documents/LIPII_Report_on_BaselineFindings.pdf`.

Granger, C. W. J., & Pesaran, M. H. (2000). Economic and Statistical Measures of Forecast Accuracy, Journal of Forecasting, 19, 537–560.

Hatch, J. K., & Frederick, L. (1998). Poverty Assessment by Microfinance Institutions: A Review of Current Practice, Microenterprise Best Practices, Development Alternatives, http://www.povertytools.org/documents/PAbyMFIs.pdf.

Henry, C., Sharma, M., Lapenu, C., & Zeller, M. (2003). Microfinance Poverty Assessment Tool, Technical Tool No. 5, Washington, D.C.: Consultative Group to Assist the Poorest, http://www.cgap.org/docs/TechnicalTool_05.pdf.

Kolesar, P., & Showers, J. L. (1985). A Robust Credit Screening Model Using Categorical Data, Management Science, 31(2), 123–133.

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

Matul, M., & Kline, S. (2003). Scoring Change: Prizma's Approach to Assessing Poverty, Spotlight Note No. 4, Warsaw: Microfinance Centre for Central and Eastern Europe and the New Independent States, http://www.mfc.org.pl/doc/Research/ImpAct/SN/MFC_SN04_eng.pdf.

Mays, E. (2000). "The Basics of Scorecard Development and Validation", in E. Mays (Ed.), Handbook of Credit Scoring (pp. 89–105). Chicago: Glenlake.

MkNelly, B., Valadez, J., Treiber, J., & Davis, R. (2001). Considering the Applicability of Lot Quality Assurance Sampling (LQAS) to Credit with Education Progress Tracking, Davis, CA: Freedom from Hunger.

Myers, J. H., & Forgy, E. W. (1963). The Development of Numerical Credit Evaluation Systems, Journal of the American Statistical Association, 58, 779–806.

Prism Research. (2003). Poverty Assessment of Prizma Clients Using the CGAP Index, Sarajevo.

Pritchett, L. (2002). It Pays to be Ignorant: A Simple Political Economy of Rigorous Program Evaluation, Journal of Policy Reform, 5(4), 251–269.

Stillwell, W. G.; Barron, F. H., & Edwards, W. (1983). Evaluating Credit Applications: A Validation of Multi-Attribute Utility Weight Elicitation Techniques, Organizational Behavior and Human Performance, 32, 87–108.

Wainer, H. (1976). Estimating Coefficients in Linear Models: It Don't Make No Nevermind, Psychological Bulletin, 83, 213–217.

World Bank; State Agency for Statistics in Bosnia-Herzegovina; Federation Statistical Institute; and the Republika Srpska Statistical Institute. (2002). Welfare in Bosnia and Herzegovina, 2001: Measurement and Findings, http://www.rzs.rs.ba/anketeLSMS/WelfareBiH.zip.

Zeller, M. (2004). Review of Poverty Assessment Tools, Accelerated Microenterprise Advancement Project, IRIS Center, University of Maryland, http://www.povertytools.org/documents/Review%20of%20Poverty%20Assessment%20Tools.pdf.

Zeller, M, Alcaraz, G.V., & Johannsen, J. (2004). Developing and Testing Poverty-Assessment Tools: Results from Accuracy Tests in Bangladesh, Accelerated Microenterprise Advancement Project, IRIS Center, University of Maryland, http://www.povertytools.org/documents/Bangladesh%20Accuracy%20Report.pdf.

NOTES

[1] While this paper was in its final stages of revision, Zeller, Alcaraz, and Johannsen (2004) appeared with an analysis of the power of poverty scorecards in Bangladesh.

[2] The poverty line in Convertible Marks per year was changed to Purchasing Power Parity dollars per day as follows. First, the ratio of PPP dollars per dollar (5.08) was derived as 2001 GDP per capita in PPP dollars (5,970) divided by GDP per capita in nominal dollars (1,175) as reported in the *2003 Human Development Report*. Second, the December 31, 2001 exchange rate of 2.22 Convertible Marks per dollar was used to convert 2,200 Convertible Marks to 991 dollars. Third, multiplying 991 dollars by the ratio of PPP dollars to dollars (5.08) gives a poverty line of 5,034 PPP dollars per year. Converting from years to days gives a poverty line of 13.79 PPP dollars per day.

[3] Statistical measures are inferior to measures based on the benefits or costs of correctly or incorrectly identifying a poor client (Granger and Pesaran, 2000). If benefits and costs are known, power curves give the information needed to evaluate scorecards.

[4] Mays (2000) says a maximum KS from 0.41 to 0.50 is “good”, 0.51 to 0.60 is “very good”, and 0.61 to 0.70 is “excellent”.

[5] By measuring poverty, Prizma risks “looking bad” vis-à-vis competitors without such measurements and who can claim (because there is no evidence to the contrary) that they have greater poverty outreach (Pritchett, 2002).

Figure 1: Correlation of indicators with poverty status, national survey

		National survey	
Indicator	**Value**	**% cases with value**	**% with value who are poor**
1. Ownership of car	No	55	26
	Yes	45	11
2. Education level of female household head/partner/spouse	≤ Primary	64	24
	> Primary	36	11
3. Number of household members	6 or more	7	45
	5	11	32
	4	26	18
	3	11	11
	2	24	6
	1	12	2
4. Ownership of stereo CD player	No	78	23
	Yes	22	8
5. Location of residence	Rural or peri-urban	75	21
	Urban	25	13
6. Average times eats meat each week with main meal	Rarely (0-2)	25	42
	Sometimes (3-5)	40	19
	Often (6 or more)	35	4
7. Average times eats sweets each week with main meal	Rarely (0-2)	47	28
	Sometimes (3-5)	31	17
	Often (6 or more)	22	5

Note: In the national survey, 19.3 percent of all cases were poor.

Figure 2: Prizma's poverty scorecard

Indicator	Value	Weight: 7 indicators	Weight: 4 indicators
1. Ownership of car	No	0	0
	Yes	12	21
2. Education level of female household head/partner/spouse	$\leq$ Primary	0	0
	$>$ Primary	4	8
3. Number of household members	6 or more	0	0
	5	8	15
	4	11	20
	3	19	32
	2	27	45
	1	34	57
4. Ownership of stereo CD player	No	0	0
	Yes	8	14
5. Location of residence	Rural or peri-urban	0	N/A
	Urban	6	N/A
6. Average times eats meat each week with main meal	Rarely (0-2)	0	N/A
	Sometimes (3-5)	8	N/A
	Often (6 or more)	20	N/A
7. Average times eats sweets each week with main meal	Rarely (0-2)	0	N/A
	Sometimes (3-5)	8	N/A
	Often (6 or more)	16	N/A
Minimum possible score (most-likely poor)		0	0
Maximum possible score (least-likely poor)		100	100

Figure 3: Surveyed households by score

Score	Cases	% of cases	Likelihood poor (%) (% with score poor in survey)
0–3	46.2	0.9	95.3
4–5	0.6	0.0	100.0
6–7	12.3	0.2	100.0
8–9	109.0	2.2	87.8
10	3.7	0.1	100.0
11	62.8	1.2	88.4
12–13	46.1	0.9	66.3
14	32.4	0.6	64.5
15	18.8	0.4	52.2
16	134.3	2.7	65.0
17	26.3	0.5	63.3
18	19.0	0.4	52.5
19	127.8	2.5	48.4
20	126.5	2.5	78.3
21	14.5	0.3	41.8
22	19.4	0.4	45.6
23	70.8	1.4	29.7
24	93.9	1.9	20.8
25	40.3	0.8	30.4
26	40.7	0.8	33.2
27	207.7	4.1	30.3
28	223.3	4.4	25.5
29	47.9	1.0	22.0
30	25.4	0.5	10.9
31	174.0	3.5	26.3
32	149.0	3.0	9.6
33	73.8	1.5	26.9
34	39.4	0.8	23.0
35	152.9	3.0	12.9
36	75.4	1.5	18.0
37	85.2	1.7	14.1
38	26.7	0.5	10.3
39	240.0	4.8	9.5
40	194.1	3.9	7.0
41	95.5	1.9	4.6
42	43.2	0.9	2.4
43	154.7	3.1	4.1
44	67.8	1.3	3.9
45	80.7	1.6	1.9
46	23.2	0.5	3.4
47	228.0	4.5	1.8
48	99.2	2.0	8.0
49	73.5	1.5	0.7
50	15.1	0.3	0.0
51	182.7	3.6	0.5
52	58.8	1.2	3.9
53	75.1	1.5	4.6
54	74.1	1.5	0.4
55	145.9	2.9	0.9
56	15.5	0.3	0.0
57	88.1	1.7	0.2
58	17.2	0.3	0.0
59	140.4	2.8	0.0
60	31.7	0.6	0.0
61	51.6	1.0	0.0
62	14.0	0.3	0.0
63	135.7	2.7	0.1
64	14.9	0.3	0.0
65	57.9	1.1	0.3
66	2.8	0.1	0.0
67	64.1	1.3	0.0
68	1.4	0.0	0.0
69	44.7	0.9	0.0
70	9.8	0.2	0.0
71	41.7	0.8	0.0
72	2.2	0.0	0.0
73	23.5	0.5	0.0
74	4.3	0.1	0.0
75	18.6	0.4	0.0
76	1.8	0.0	0.0
77	21.0	0.4	0.0
78	4.9	0.1	0.0
79	21.6	0.4	0.0
80	1.7	0.0	0.0
81–82	5.7	0.1	0.0
83–84	0.7	0.0	0.0
85	13.6	0.3	0.0
86	0.9	0.0	0.0
87	4.8	0.1	0.0
88	0.2	0.0	0.0
89–91	0.3	0.0	0.0
92	0.3	0.0	0.0
93	3.1	0.1	0.0
94–99	1.0	0.0	0.0
100	0.3	0.0	0.0
Total:	5,039.6	100.0	19.3

Figure 4: Power Curve

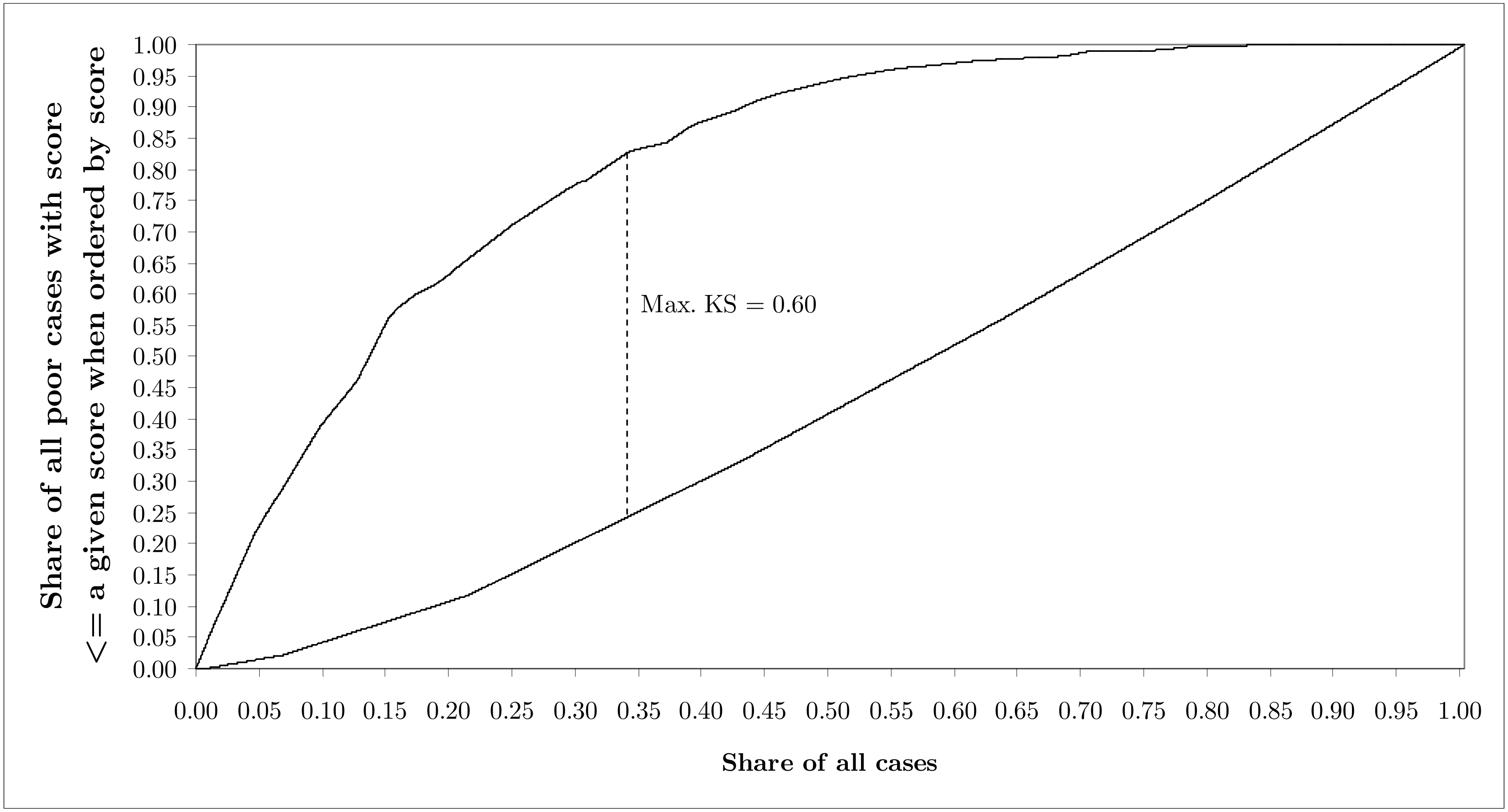

Figure 5: Importance of scorecard indicators

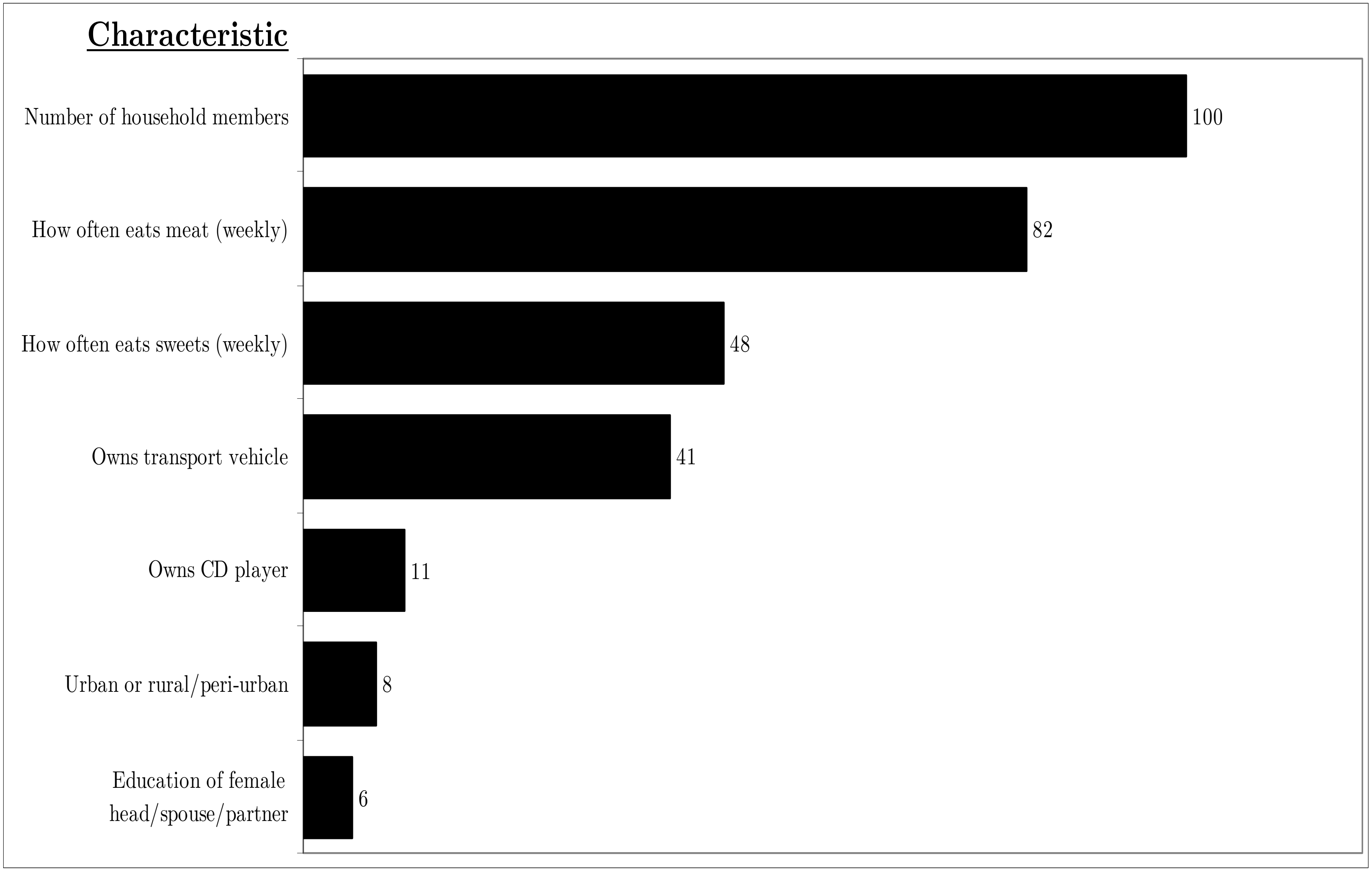

Figure 6: Overall poverty rate, new Prizma borrowers

Score	Cases	% of cases	Likelihood poor (%) (% with score poor in survey)
0–3	36	0.7	95.3
4–5	18	0.3	100.0
6–7	5	0.1	100.0
8–9	101	2.0	87.8
10	2	0.0	100.0
11	46	0.9	88.4
12–13	96	1.9	66.3
14	3	0.1	64.5
15	45	0.9	52.2
16	131	2.5	65.0
17	4	0.1	63.3
18	10	0.2	52.5
19	72	1.4	48.4
20	153	3.0	78.3
21	2	0.0	41.8
22	10	0.2	45.6
23	92	1.8	29.7
24	152	2.9	20.8
25	13	0.3	30.4
26	29	0.6	33.2
27	140	2.7	30.3
28	212	4.1	25.5
29	17	0.3	22.0
30	21	0.4	10.9
31	172	3.3	26.3
32	195	3.8	9.6
33	16	0.3	26.9
34	56	1.1	23.0
35	132	2.5	12.9
36	210	4.1	18.0
37	43	0.8	14.1
38	45	0.9	10.3
39	178	3.4	9.5
40	220	4.2	7.0
41	30	0.6	4.6
42	43	0.8	2.4
43	169	3.3	4.1
44	117	2.3	3.9
45	63	1.2	1.9
46	88	1.7	3.4
47	137	2.6	1.8
48	188	3.6	8.0
49	55	1.1	0.7
50	24	0.5	0.0
51	161	3.1	0.5
52	109	2.1	3.9
53	38	0.7	4.6
54	75	1.4	0.4
55	83	1.6	0.9
56	144	2.8	0.0
57	66	1.3	0.2
58	30	0.6	0.0
59	106	2.0	0.0
60	140	2.7	0.0
61	35	0.7	0.0
62	45	0.9	0.0
63	68	1.3	0.1
64	24	0.5	0.0
65	44	0.8	0.3
66	64	1.2	0.0
67	36	0.7	0.0
68	77	1.5	0.0
69	21	0.4	0.0
70	1	0.0	0.0
71–72	57	1.1	0.0
73	20	0.4	0.0
74	50	1.0	0.0
75–76	10	0.2	0.0
77–78	52	1.0	0.0
79–80	13	0.3	0.0
81–84	5	0.1	0.0
85–86	9	0.2	0.0
87–92	1	0.0	0.0
93–99	2	0.0	0.0
100	0	0.0	0.0
Total:	5,177	100.0	17.9

Figure 7: Overall poverty rate by branch

Branch	Cases	Poverty rate
Banja Luka	655	6.1
Mostar	745	13.0
Bihać	1,576	14.5
Zenica	998	23.9
Sarajevo	1,203	27.1
Total:	5,177	17.9

Figure 8: Overall poverty rate by loan product

Product	Cases	Share (%) loans to individuals	Poverty rate
Farming	64	100	11.1
Enterprise	2,777	0	15.9
Basic needs	2,062	91	20.4
Small farm	211	48	21.1
Housing	63	100	22.2
Total:	5,177	41	17.9

Figure 9: Power curve for the seven-indicator and fully benchmarked scorecards

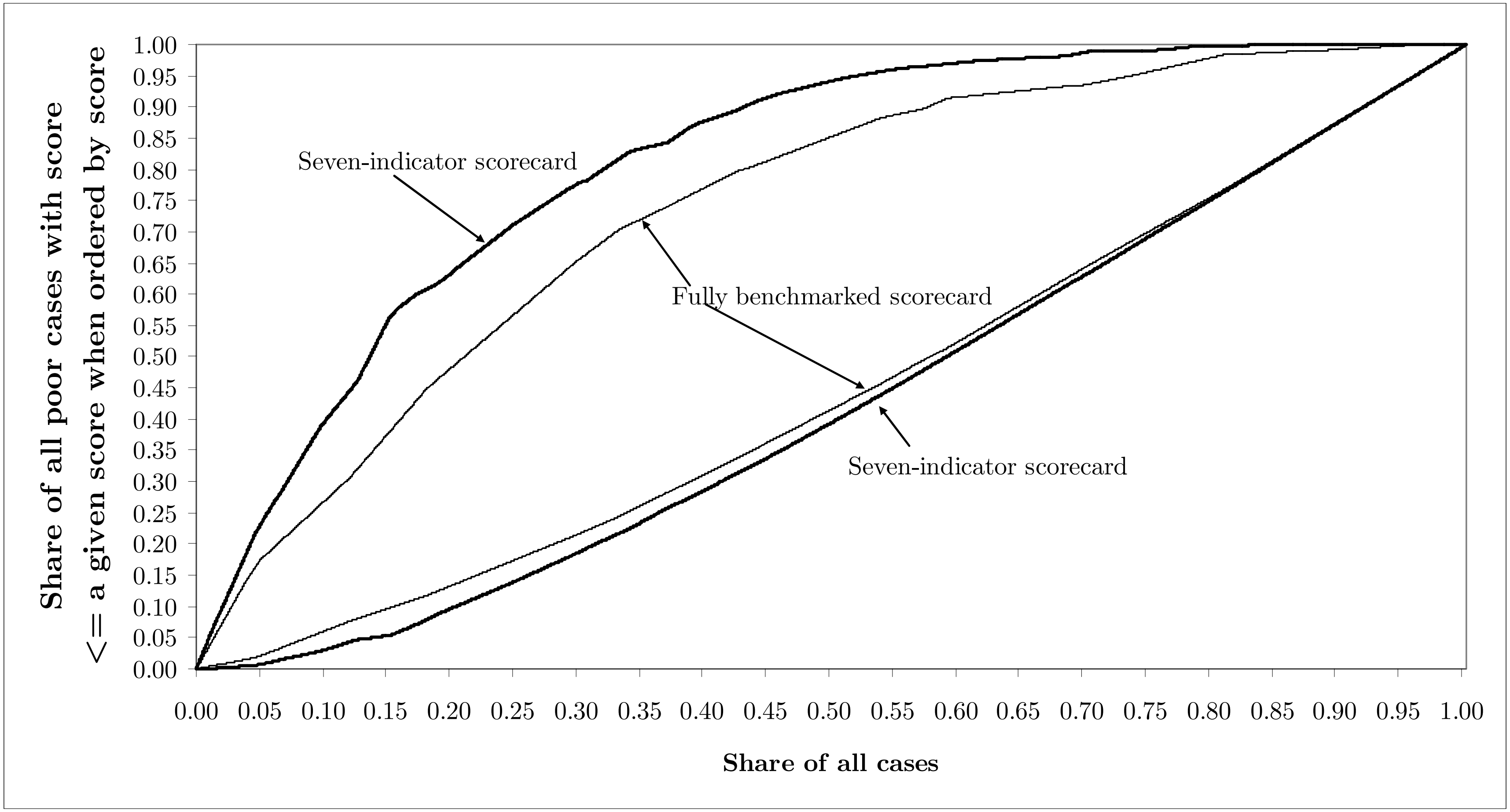

Figure 10: Prizma's overall poverty rate with a four-indicator, fully benchmarked scorecard

Score	Cases	% of cases	Likelihood poor (%) (% with score poor in survey)
0–7	204.0	3.9	69.2
8–11	195.0	3.8	58.7
12–14	188.0	3.6	36.1
15	196.0	3.8	45.4
16–19	69.0	1.3	58.3
20–22	484.0	9.3	33.5
23–26	329.0	6.4	29.3
27	182.0	3.5	18.5
28–30	488.0	9.4	19.7
31	145.0	2.8	6.4
32–34	249.0	4.8	14.9
35	484.0	9.3	8.4
36–39	142.0	2.7	18.3
40–42	474.0	9.2	3.9
43	403.0	7.8	2.9
44	71.0	1.4	7.5
45–47	89.0	1.7	8.9
48–51	369.0	7.1	2.0
52	148.0	2.9	1.3
53–56	105.0	2.0	1.8
57–59	5.0	0.1	2.8
60	94.0	1.8	0.0
61–64	36.0	0.7	0.0
65–68	18.0	0.3	0.0
69–72	8.0	0.2	0.0
73–76	2.0	0.0	0.0
77–84	0.0	0.0	0.0
85–100	0.0	0.0	0.0
Total:	5,177.0	100.0	19.5

Figure 11: Loan-size-only scorecard based on Prizma's poverty scorecard for new borrowers

Amount disbursed	Score	% of cases	Likelihood poor
0 to 400 KM	0	4.3	23.8
401 to 599 KM	1	35.6	20.0
600 to 800 KM	2	13.8	19.6
801 to 1000 KM	3	18.2	17.6
1001 KM or more	4	28.1	13.8
	Total:	100.0	17.9

Figure 12: Power curve, seven-indicator and loan-size-only scorecards