
Targeting Outcomes Redux

David Coady • Margaret Grosh • John Hoddinott

A newly constructed comprehensive database of 122 targeted antipoverty interventions in 48 countries is used to examine the contested issue of the efficacy of targeting interventions in developing countries. Though the median program transfers 25 percent more to poor individuals (those in the bottom two quintiles) than would universal allocation, a quarter of the interventions are regressive. Targeting is better in richer countries, in countries where governments are more likely to be held accountable, and in countries where inequality is higher. Interventions that use means testing, geographic targeting, and self-selection based on a work requirement are all associated with an increased share of benefits going to poor people. Proxy-means testing, community-based selection, and demographic targeting to children show good results on average but with wide variation. Self-selection based on consumption, demographic targeting to the elderly, and community bidding show limited potential for good targeting. The substantial variation in targeting performance within specific program types and specific targeting methods suggests that differences in implementation are also important factors in determining the success of targeting to poor individuals.

In the past two decades a consensus has emerged that social safety nets designed to raise and protect the consumption levels of poor households can play a crucial role in development (World Bank 1990, 1997, 2000). Implementing this agenda for reducing poverty requires methods for reaching poor people. This can be accomplished by broad targeting through programs that reach a wide swath of society, including poor people (for example, universal primary education, an extensive network of basic health care), or by narrow targeting through programs that identify the poor more specifically to confer benefits disproportionately on this group (for example, income transfer programs).¹ The case for narrow targeting rests on the existence of a budget constraint.² The overall poverty impact of a program depends on the number of poor households covered and the level of benefits they receive. With a fixed poverty alleviation budget, the opportunity cost of transfers “leaking” to nonpoor households is a lower impact on poverty reduction, reflecting less coverage

of poor households or lower benefit levels. Targeting transfers to poor households means that more benefits can go to them.

Divergent views on the efficacy of narrowly targeted interventions are based on differing assessments of three questions: Are such methods likely to achieve better targeting outcomes? Are they cost-effective? Do they raise the living standards of poor people? This article addresses the first question.³ The fairly extensive literature on this topic is dominated by descriptions of individual, sometimes idiosyncratic programs. Even comparative analyses tend to cover a single region (Grosh 1994 for Latin America and the Caribbean; Braithwaite, Grootaert, and Milanovic 2000 for Eastern Europe and Central Asia), method (Bigman and Fofack 2000 on geographic targeting), or intervention (Rawlings, Sherburne-Benz, and van Domelen 2001 on social funds). This partial coverage frustrates efforts to make broader assessments about the effectiveness of different targeting methods or to draw policy-relevant lessons.

To rectify this weakness, this article draws on a newly constructed database of 122 targeted antipoverty interventions in 48 countries in Latin America and the Caribbean, Europe and Central Asia (including the former Soviet Union), the Middle East and North Africa, Sub-Saharan Africa, and South and East Asia. These data are used to address three questions: What are the targeting outcomes? Are there systematic differences in targeting performance by targeting method or other factors? What are the implications of any such systematic differences for the design and implementation of targeted interventions?

The analysis shows that the median targeted program is progressive in that it transfers 25 percent more to poor individuals—persons and households in the bottom two income quintiles—than would be the case with a universal allocation. However, for a staggering quarter of the programs, outcomes are regressive. Countries with higher income, here taken to imply better capacity for program implementation, do better at directing benefits toward poorer members of the population, as do countries where governments are more likely to be held accountable for their behavior, as suggested by better measures of voice. Targeting is also better in countries where inequality is more pronounced. This is consistent with higher welfare gains from targeting when inequality is high as well as lower costs in identifying poor beneficiaries.

Subject to several caveats, interventions that use means testing, geographic targeting, and self-selection based on a work requirement are all associated with an increased share of benefits going to the bottom two quintiles. Proxy-means testing, community-based selection of individuals, and demographic targeting to children show good results on average but with wide variation. Self-selection based on consumption, demographic targeting to the elderly, and community bidding show limited potential for good targeting. That said, examination of experiences with specific program types and specific targeting methods shows considerable variation in targeting performance. This partly reflects heterogeneity across the interventions assessed—some, like means-tested transfers, have poverty reduction as their sole objective,

whereas others, such as social funds, have multiple objectives. However, substantial variation in targeting performance by specific program types and specific targeting methods suggests that differences in implementation are also important factors in whether targeting to poor individuals is successful or not.

Data Construction and Description

The first step in the analysis was a review of the extensive literature on targeted programs and the construction of a database of targeted antipoverty interventions.⁴ To our knowledge, this work represents the most extensive attempt to construct such a database.

Database Construction

There were five criteria for inclusion in this database:

- The intervention is in a low- or middle-income country.
- A principal objective of the intervention is poverty reduction, defined in terms of income or consumption.
- The intervention involves a transfer of resources to the beneficiary.
- Documentation on the intervention contains information on the type of targeting method, its implementation, and outcomes.
- The intervention is relatively recent (generally during 1985–2003).

Included in the data are cash transfers (including welfare and social assistance payments, child benefits, and noncontributory pensions), near-cash transfers (such as quantity-rationed subsidized food and food stamps), food transfers, universal food subsidies, nonfood subsidies, public works, and social funds.

Two observations should be made on these criteria for inclusion. First, some interventions have broad objectives that may go beyond direct poverty reduction. Social funds are a good example. Although short-term poverty reduction can be an important component of these interventions, so can be the construction of physical assets valued by the poor and the development of local capacity to design, implement, and maintain infrastructure. The heterogeneity of objectives within broadly defined “antipoverty” interventions demands caution in interpreting comparisons across types of interventions.

Second, focusing the review in this way necessarily means excluding some interventions that may be targeted and may have some impact on poverty. Thus, excluded are occupationally based transfer schemes, such as formal sector unemployment insurance or occupational old age or disability pensions (the principal mechanism determining eligibility and benefit levels are employment and contributions history rather than poverty status); credit and microcredit schemes (although often targeted,

they are largely motivated by credit market failures and do not necessarily involve a transfer of resources to participants);⁵ supplementary feeding programs (mainly because the vast literature on this type of intervention did not yield studies that satisfied the criteria described); and most short-term emergency aid (because despite the clear poverty focus and targeting by need, the time scale on which it operates typically precludes an assessment of the distribution of the benefits).

Because most studies of targeting do not appear in peer-reviewed journals, searches were conducted of the gray literature using Web search engines of the World Bank, Eldis, and the International Food Policy Research Institute using the following key words: *safety nets*, *targeting*, *social funds*, *pensions*, *public works*, and *subsidies*. Additional studies were found by canvassing colleagues about work that had not yet been catalogued. Searches were also undertaken in the following academic journals for 1990–2003: *Economic Development and Cultural Change*, *Journal of Development Economics*, *Journal of Development Studies*, *Journal of Public Economics*, *World Bank Economic Review*, *World Bank Research Observer*, *World Development*, and *Economic and Political Weekly*. Existing compilations, such as Grosh (1994) and Braithwaite, Grootaert, and Milanovic (2000), were also reviewed.

The sample of interventions is not necessarily reflective of the distribution of programs around the world but rather of programs that have some measurement of targeting outcomes and that have been written up in the catalogued English language literature that was accessed in this study. A program is more likely to be written up this way if one or more of the following features apply: It is from a country with a household survey that measures consumption and participation in government programs, it is in a country with a culture of evaluation as part of decision-making, it receives funding from an international agency that requires measurement of outcomes, and it appeals to analysts and editors because of its methods or setting. This suggests, for example, that programs using community-based methods and agents are underrepresented. Community-based methods are often locally funded and chosen when data and administrative capacity are weak, features that reduce the likelihood of an evaluation being done and finding its way into the international literature. The literature on public works in Sub-Saharan Africa is probably also underrepresented. By contrast, proxy-means tests are well represented, with a large share of all such programs showing up in the sample.

Database Description

Based on the criteria described, information was collected on 122 interventions in 48 countries.

Intervention type. The distribution of these interventions shows fairly broad coverage by region and by intervention type (table 1). In some regions a particular intervention type dominates: cash transfers in Europe and Central Asia, universal food subsidies in

Table 1. Distribution of Intervention Types by Region and Country Income Level for Sample Programs in 1985–2003

<i>Region and income level^a</i>	<i>Transfers</i>			<i>Subsidies</i>		<i>Public Works</i>	
	<i>Cash</i>	<i>Near Cash^b</i>	<i>Food</i>	<i>Food</i>	<i>Nonfood</i>	<i>Job Creation</i>	<i>Program Output^c</i>
<i>Region</i>							
Latin America and Caribbean (32)	14	3	4	0	2	4	5
Europe and Central Asia (26)	24	1	0	0	0	0	1
Middle East and North Africa (14)	1	0	0	13	0	0	0
Sub-Saharan Africa (13)	4	0	1	4	1	2	1
South Asia (21)	1	13	3	0	0	4	0
East Asia and Pacific (16)	5	1	4	1	2	2	1
<i>Income level^d</i>							
Poorest (63)	17	15	6	10	2	8	5
Less poor (59)	32	3	6	8	3	4	3
Total (122)	49	18	12	18	5	12	8

^aNumbers in parentheses are total number of interventions.

^bIncludes interventions such as food stamps or the right to purchase a limited quantity of food at a subsidized price.

^cFor example, social funds.

^dPoorest countries have per capita GDP in 1995 purchasing power parity of \$1,200 or less; less poor countries have per capita GDP of between \$1,201 and \$10,840.

Source: Authors' compilation based on data search described in the text.

the Middle East and North Africa, and near-cash transfers in South Asia. There is a wider mix of reported interventions in Latin American and the Caribbean, Sub-Saharan Africa, and East Asia and Pacific. Most of the cash transfer programs and social funds are in Latin America and the Caribbean and Europe and Central Asia, most of the near-cash transfer programs are in South Asia, most of the universal food subsidies are in the Middle East and North Africa, and most of the social funds are in Latin America and the Caribbean. Cash transfer programs are more likely to be found in less poor countries and near-cash transfers in the poorest countries. Cash transfer programs account for the largest proportion (40 percent) of intervention types.

Targeting forms. The database distinguishes three broad forms of targeting: individual and household assessments, categorical (geography, age), and self-selection, each with subcategories (table 2).

With *individual and household assessments*, eligibility is directly assessed on an individual basis using one of several methods. In a *verified means test*, information is obtained on a household's income or wealth and compared with other sources of information, such as pay stubs and income and property tax records. Use of this

Table 2. Distribution of Targeting Forms and Methods by Region, Country Income Levels, and Program Type for Sample Programs in 1985–2003

<i>Region and income level^a</i>	<i>Individual assessment</i>			<i>Categorical</i>				<i>Self-selection</i>		
	<i>Means tests</i>	<i>Proxy-means tests</i>	<i>Community assessment</i>	<i>Geography</i>	<i>Age, elderly</i>	<i>Age, children</i>	<i>Other</i>	<i>Work</i>	<i>Consumption</i>	<i>Community bidding</i>
<i>Region</i>										
Latin America and Caribbean (68)	8	5	3	20	4	14	4	4	0	6
Europe and Central Asia (46)	14	1	3	1	6	11	7	2	0	1
Middle East & North Africa (23)	4	0	0	2	1	1	2	0	12	1
Sub-Saharan Africa (25)	3	0	2	3	5	1	4	2	4	1
South Asia (49)	2	1	3	16	2	1	10	4	10	0
East Asia (42)	3	1	3	10	6	8	8	1	1	1
<i>Income level^b</i>										
Poorest (147)	12	3	10	37	10	14	28	8	19	6
Less poor (106)	22	5	4	15	14	22	7	5	8	4
<i>Program type</i>										
Cash transfer (103)	24	4	5	9	19	24	16	2	0	0
Near-cash transfer (36)	4	3	0	12	1	2	4	0	10	0
Food transfer (35)	0	1	5	9	3	9	7	0	0	1
Food subsidy (23)	3	0	0	2	0	0	0	0	17	1
Nonfood subsidy (9)	3	0	0	2	1	1	2	0	0	0
Public works, job creation (29)	0	0	2	10	0	0	6	11	0	0
Public works, program output (e.g., social fund) (18)	0	0	2	8	0	0	0	0	0	8
Total (253)	34	8	14	52	24	36	35	13	27	10

^aNumbers in parentheses are total number of interventions. Many programs use more than one targeting method, so the total number of targeting methods is greater than the number of programs.

^bPoorest countries have per capita GDP in 1995 purchasing power parity of \$1 200 or less; less poor countries have per capita GDP of between \$1,201 and \$10,840.

Source: Authors' compilation based on data search described in the text.

method requires the existence of verifiable records in the target population and the administrative capacity to process the information and update it regularly.

A *simple means test*, with no independent verification of income, can be conducted by visiting the household to verify qualitatively that visible standards of living (reflecting income or wealth) are generally consistent with reported figures. *Proxy-means tests* involve generating a score for applicants based on fairly easy to observe characteristics of the household, such as location and quality of dwelling, ownership of durable goods, demographic structure of the household, and the education of adult members. The indicators used in calculating this score and their weights are derived from statistical analysis of data from detailed household surveys.

An increasingly popular approach to individual assessment has been to decentralize the selection process to local communities. A group of community members or a community leader whose principal functions in the community are not related to the transfer program then decides whom in the community should benefit and whom should not—hence the term *community-based targeting*.

Categorical targeting, also referred to as statistical targeting, tagging, or group targeting, involves defining eligibility in terms of individual or household characteristics that are considered easy to observe, hard to falsely manipulate, and correlated with poverty. Age, gender, ethnicity, land ownership, and household demographic composition or location are common methods of targeting. Geographic targeting is often used in tandem with other methods.

Some interventions rely on *self-selection*. These interventions, with nominally universal eligibility, are designed with dimensions intended to encourage the poorest to use the program and the nonpoor not to do so.⁶ This is accomplished by recognizing differences in the private participation costs of poor and nonpoor households. Examples include the use of low wages on *public works schemes* so that only those with a low opportunity cost of time due to low wages or limited hours of employment will present themselves for work; time restrictions on transfers, requiring applicants to queue; and location of points of service delivery in areas where the poor are highly concentrated so that the nonpoor have higher (private and social) costs of access. In social fund-type interventions self-selection occurs at the community level (*community bidding*), with targeting occurring as a result of differences in private participation costs between poor and nonpoor communities.⁷

Universal food subsidies (with or without quantity rationing) can be viewed as a form of self-selection because they are universally available and households receive benefits by deciding to consume the commodity. Households can often determine not just whether to participate but also the intensity of their participation. The more income elastic are the expenditures on these items, the more effective is the targeting. Food transfers often involve commodities with “inferior” characteristics (for example, low-quality wheat or rice), and households often substitute away from such commodities as their incomes rise.⁸

Targeting methods. Within this broad taxonomy of targeting forms, interventions rely on various targeting methods and often on several methods (see table 2). The 122 interventions used 253 different targeting methods, for an average of just over two targeting methods per intervention. Although 48 interventions used a single targeting method, 42 used two methods, 21 used three methods, and 11 used four or more methods.

Thus, although certain intervention types typically rely on specific targeting methods, most use a combination of methods, presumably because there is a synergy between them that results in greater targeting efficiency. Public works programs typically use a combination of geographic targeting and self-selection based on low wages and a work requirement. If demand exceeds supply at the wage paid, public works also often require additional rationing of employment using categorical targeting. Similarly, social funds are partly demand driven and therefore have an element of community self-selection. Food subsidies are self-targeted based on consumption patterns. Cash transfers are most likely to have some form of individual assessment but are also often conditioned on other characteristics (such as age in the case of pensions or child benefit).

There are some marked differences in targeting methods by region. Most interventions using means and proxy-means testing are concentrated in Europe and Central Asia and Latin America and the Caribbean. A legacy of the central planning era in Europe and Central Asia is an extensive administrative system suited to individual assessment using some form of means or proxy-means testing. This, together with a distribution of income that at least at the time of transition was relatively equal, has meant that targeting is based either on individual assessment or some individual characteristic, such as age. A reliance on food subsidies explains why self-targeting based on consumption patterns is the dominant targeting method in the Middle East and North Africa. South Asia is notable for extensive use of geographic targeting and for relatively high reliance on self-selection based on work or consumption. Latin American and Caribbean countries also use geographic targeting extensively, along with means or proxy-means testing or targeting to children. Interventions in Sub-Saharan Africa and East Asia show more mixed patterns. There are also broad differences across income levels. Generally, poorer countries rely more on self-selection and categorical targeting, whereas individual assessments are relatively more common in less poor countries. The one exception to these general patterns is categorical targeting by age, which is used relatively less frequently in poor countries.

Assessing Targeting Effectiveness

This section outlines the methodology used to compare targeting performance across interventions by creating an indicator of targeting performance. It also identifies some important caveats to be kept in mind when interpreting this indicator and briefly describes targeting outcomes in terms of this indicator of targeting performance.

Measures of Targeting Effectiveness

There is a growing body of literature on measuring targeting effectiveness (Ravallion and Chao 1989; Ravallion 1993; Cornia and Stewart 1995; Coady and Skoufias 2001). Comparing the performance of the targeting methods used in the range of programs considered in this meta-analysis requires a comparable performance indicator for each program. The definitions, methods, and presentations in the original studies vary in ways that make it difficult to assemble a single summary performance indicator. Incidence and participation rates may be reported over the full welfare distribution; for the poorest 10, 20, or 40 percent of the population; or for a poor and nonpoor classification that differs by country. Other studies report none of these measures but use other, less common ones. Of course, the measures of welfare used are not always strictly comparable. The problem is how best to compare targeting performance outcomes using data that are not strictly comparable.

Most studies catalogued in the database provide information on at least one of the following indices:

- The proportion of total transfers received by individuals or households falling within the bottom 40, 20, or 10 percent of the national income distribution.⁹
- The proportion of beneficiaries falling within the bottom 40, 20, or 10 percent of the national income distribution.
- The proportion of total transfers going to poor households or individuals, defined in terms of some specified part of the welfare distribution (for example, falling in the bottom 35 percent of the income distribution).

Ideally, information would be available on the proportion of total transfers received by households falling within different deciles (40th, 20th, 10th, and so on) of the national income distribution. This is a better measure than the proportion of beneficiaries by decile, which says nothing about variations in the levels of transfers. These two measures—proportion of total transfers and proportion of beneficiaries—are only equivalent when transfer levels are uniform across beneficiaries.

Because the studies did not use a single common measure of targeting performance, a uniform measure of targeting effectiveness was constructed based on a comparison of actual performance and a common reference outcome: the outcome that would result from neutral (as opposed to progressive or regressive) targeting. A neutral targeting outcome means that each decile receives 10 percent of the transfer budget or that each decile accounts for 10 percent of the program beneficiaries. Neutral targeting can be thought to arise either from the random allocation of benefits across the population or from a universal intervention in which all individuals receive identical benefits.

The indicator is constructed by dividing the actual outcome by the appropriate neutral outcome. For example, if people or households in the bottom 40 percent of

the income distribution receive 60 percent of the benefits, the indicator of performance is calculated as $60/40 = 1.5$, meaning that targeting has resulted in the target group (here, people in the bottom two quintiles) receiving 50 percent more than they would have received under a universal intervention. A value greater than one indicates progressive targeting, and a value less than one regressive targeting (unity denotes neutral targeting).

The performance indicator used in the analysis that follows is based on a lexicographic selection process among the available incidence indicators, as follows. Preference is given to a measure of the share of benefits accruing to a target group rather than to a measure based on a portion of beneficiaries in the target group. Within that, preference is giving to the proportion accruing to the bottom two quintiles. If that information is not available, then to the bottom quintile, and if that is not available, then to the poorest decile. Such a performance indicator could be calculated for 85 of the 122 programs (table 3).

Descriptive Results

There is enormous variation in targeting performance, with scores ranging from 4 for the Trabajar public works program in Argentina to 0.28 for value-added tax exemptions on fresh milk in South Africa. The median value is 1.25, so that the “typical” program transfers 25 percent more to households in the bottom quintiles or below the poverty line than would be the case with a universal allocation. However, a staggering 21 of the 85 programs—25 percent—are regressive, with a performance index of less than 1, suggesting that in these cases a random selection of beneficiaries would provide greater benefits to the poor.

It is instructive to focus on the 10 worst and 10 best programs. The 10 worst have scores ranging from 0.28 to 0.78 with a median score of 0.60. Five are food subsidy programs and three involve cash transfers. Median performance rises from 1.25 to 1.3 if interventions using self-selection based on consumption are withdrawn from the sample, and the proportion of regressive interventions drops from 25 percent to 16 percent. Only one of the poorly performing programs uses either means or proxy-means targeting methods, none use geographic targeting, and the countries come from across the income spectrum. Scores for the top 10 programs range from 2.02 to 4.00, with a median score of 2.15, and all but one are in either Latin America and the Caribbean or Europe and Central Asia. Of these 10, 9 involve cash transfers; 9 use means, proxy-means, or geographic targeting; and 7 are in less poor countries.

Cash transfers show up in both the 10 best and 10 worst programs, highlighting the possibility that variations in targeting performance may reflect poor implementation rather than poor program potential. However, public works programs are all in the top half of the performance table, and social funds are nearly all in the bottom half. This is consistent with there being a tradeoff between the objective of reducing

Table 3. Ranking of Targeting Performance of Antipoverty Interventions for Sample Programs in 1985–2003

Country	Program	Individual assessment			Categorical			Self-selection		Share of Transferees going to		
		Performance	Means		Geographic	Age		Consumption	Community bidding	Poorest 10%	Poorest 20%	Poorest 40%
			test	Proxy-means test		Elderly	Children					
Argentina	Trabajar public works	4.00			✓			✓		60.0	80.0	
Estonia	Cash social assistance	3.47	✓							34.7		
Dominica	Cash transfers	3.00	✓								60.0	
Hungary	Cash social assistance	2.72	✓							27.2		
Albania	Ngjime Ekonomika cash	2.65		✓				✓		36.0	53.0	
Yemen	Social welfare fund cash	2.15	✓		✓	✓				68.0	75.0	86.0
Poland	Social assistance cash	2.10	✓					✓		21.0		
Romania	Minimum income guarantee	2.08	✓					✓			62.0	83.0
Chile	SUF cash transfers	2.08		✓			✓				57.0	83.0
Nicaragua	RPS conditional cash transfer	2.02			✓		✓			32.6	55.0	80.9
Honduras	PRAF cash transfer	1.99					✓			22.1	42.5	79.5
Chile	School lunches	1.98		✓	✓		✓				53.0	79.0
Slovenia	Otroski Dodatek child benefit	1.95	✓		✓		✓			42.0	65.0	78.0
Bulgaria	Cash social assistance	1.95	✓							22.0	60.5	78.1
Bolivia	ESF public works	1.93			✓			✓		25.0	31.0	77.0
Kyrgyz Republic	Unified monthly cash benefit	1.85									54.0	74.0
Chile	PASIS cash to poor elderly	1.83		✓		✓					50.0	73.0
Republic of Korea	LPS cash for home care	1.81	✓			✓				32.3	56.8	72.3
Peru	“Glass of milk” program	1.80					✓		✓			
Chile	PEM/POJH public works	1.78			✓			✓			42.0	72.0
Costa Rica	Noncontributory pensions	1.73	✓			✓					51.0	71.0
								✓			45.0	69.0
Indonesia	JPS-BK health subsidy	1.68			✓						33.6	
India	Rural AP State PDS rice	1.63	✓		✓			✓		19.1	35.8	65.2
Costa Rica	School Lunches	1.63			✓		✓				39.0	65.0

(Continued)

Table 3. (Continued)

Country	Program	Individual assessment				Categorical			Self-selection			Share of Transferees going to			
		Performance	Means test	Proxy-means test	Community assessment	Geographic	Age		Other	Work	Consumption	Community bidding	Poorest 10%	Poorest 20%	Poorest 40%
							Elderly	Children							
Mexico	LICONSA milk subsidy	1.60	✓			✓		✓							64.0
India	PDS-subsidized jowar	1.58				✓					✓				63.0
Hungary	Child cash allowance	1.57						✓					23.5	39.5	62.6
Mexico	PROGRESA conditional cash	1.56		✓		✓		✓					22.0	39.5	62.4
Colombia	Housing subsidy	1.50	✓											35.0	60.0
Colombia	ICBF day care/school food	1.50		✓				✓						28.0	60.0
Indonesia	JPS-PDK public works	1.48							✓	✓				29.6	
Jamaica	Food stamps – MCH	1.45				✓		✓	✓					29.0	58.0
Indonesia	JPS education subsidy	1.44			✓	✓		✓	✓					28.8	
India	Rural Mh. State PDS	1.36				✓					✓		10.4	26.3	54.5
Zambia	Housing subsidy	1.35	✓											28.0	54.0
Uzbekistan	Child allowance	1.35	✓		✓			✓	✓					28.4	54.0
Latvia	Family child allowance	1.33						✓			✓		15.0	29.0	53.0
India	PDS-subsidized rice	1.33				✓									53.0
Indonesia	JPS-OPK rice subsidy	1.32		✓		✓			✓					26.4	
Bolivia	Social investment fund	1.30				✓			✓			✓	11.0	24.5	52.0
Jamaica	Food stamps program	1.30	✓					✓	✓					31.0	52.0
Honduras	Social fund program	1.25				✓						✓	19.0	32.0	50.0
Chile	Viviendas Basicas mortgage	1.25		✓										28.0	50.0
India	PDS-subsidized oil	1.25				✓					✓				50.0
Sri Lanka	Food stamps	1.25	✓											28.0	50.0
South Africa	Maize VAT exemption	1.23									✓				
Vietnam	SGF food transfers	1.22				✓		✓	✓					24.4	48.0
India	PDS-subsidized kerosene	1.20				✓					✓				
Bangladesh	Food for education	1.20			✓	✓		✓	✓						
Morocco	Food subsidies, flour	1.18									✓			23.0	47.0
India	PDS-subsidized sugar	1.13				✓					✓			45.0	45.0
Armenia	Family cash benefit	1.13						✓					14.0	25.0	45.0
Peru	Social fund program	1.10				✓						✓		23.0	44.0

current poverty (through public works wage transfers) and the objective of reducing future poverty through developmental public investments (through the assets created by social fund programs). Also, the predominance of less poor countries among the top half of the table suggests that characteristics correlated with income, such as administrative capacity, are important to targeting performance.

Summary statistics on targeting performance—sample size, median, inter-quartile range, and the inter-quartile range as a percentage of the median—by targeting form and method initially suggest a clear hierarchy of targeting performance (table 4). As might be expected, interventions using individual assessments outperform interventions relying on categorical targeting, which in turn outperform interventions relying on self-selection.

Closer inspection, however, reveals that such impressions are too general to be useful. First, there is considerable heterogeneity within these broad methods of targeting. Most notably, self-selection includes interventions that have the highest median performance (those using a work requirement) and those that have the lowest median performance (self-selection based on consumption). Second, self-selection interventions based on consumption and community bidding for interventions have lower median values than other interventions and relatively low variations in these values as measured by the inter-quartile range as a percentage of the median. This

Table 4. Targeting Performance by Targeting Form and Method for Sample Programs in 1985–2003

<i>Targeting form and method</i>	<i>Sample size</i>	<i>Median targeting performance</i>	<i>Interquartile range</i>	<i>Interquartile range as % of median</i>
<i>All methods</i>	85	1.25	0.68	54.4
<i>Any individual assessment</i>	37	1.50	0.75	50.0
Means testing	26	1.55	0.90	58.1
Proxy-means testing	7	1.50	0.58	38.7
Community assessment	6	1.40	0.78	55.7
<i>Any categorical method</i>	58	1.32	0.64	48.5
Geographic	33	1.33	0.51	36.9
Age, elderly	12	1.16	0.81	69.8
Age, young	26	1.53	0.65	42.5
Other categorical	17	1.35	0.48	35.6
<i>Any self-selection method</i>	38	1.10	0.41	37.2
Work	6	1.89	0.30	15.9
Consumption	25	1.00	0.35	35.0
Community bidding	7	1.10	0.22	20.0

Source: Authors' compilation based on data search described in the text.

suggests that other things being equal, even the best examples of these targeting methods produce relatively small targeting gains. By contrast, whereas other methods report higher median values, they also tend to have proportionately higher variations in targeting effectiveness. So although these methods offer potentially large gains, there is no guarantee they will improve targeting performance.

One way to explore the source of variation in targeting outcomes is with a Theil inequality index. An attractive feature of the Theil index is that it is subgroup decomposable. When the data are grouped by some characteristic, variations in targeting can be allocated across these programs into two categories: targeting variations due to variations within each group and targeting variations due to variations across groups. When programs are grouped by region, variations in average performance across continents explain only about 28 percent of total variation. When programs are grouped by type, variations in average performance between programs explain 36 percent of the total variation. Grouping by targeting method (geographic, means or proxy-means, both, or other targeting methods) explains only 20 percent of the total variation.

One way of interpreting these large variations is to consider implementation effectiveness. No matter how good the choice among methods or programs, effectiveness of implementation is a key factor determining targeting performance. This point is further strengthened by noting that raising the performance of all programs with the same targeting method and with performance below the method median to the median for that method increases mean targeting performance from 1.38 to 1.55, a return of 17 percentage points. This issue is picked up again in the section on the regression analysis.

Caveats and Limitations

Before reporting on the regression analysis, it is important to make explicit several caveats for interpreting the performance measure and, thus, the analysis based on it.

First, the performance measure is a combination of various measures as already discussed, although the measure used for the vast majority of the interventions (80 percent) is the percentage of benefits accruing to the bottom 40 percent or 20 percent of the national income distribution. This raises concerns about comparability. For example, it may be more difficult to target the poorest 20 percent than the poorest 40 percent, so programs assessed using the 20 percent measure may appear relatively ineffective solely because of the performance indicator used.

This issue was addressed in a number of ways. A second performance measure was calculated that gives, through its lexicographic ordering, priority to the proportion of resources flowing to the bottom 10 percent, then the bottom 20 percent, then the bottom 40 percent. There are no meaningful changes to the results reported in tables 3 and 4 using this performance measure. All regressions were also run

(reported later) using both measures of targeting performance, and again there were no meaningful changes to the results. This is not completely surprising given that the performance measure and the alternative have correlation coefficients of 0.94 (for levels) and 0.97 (for ranks). As a further check, the multivariate regressions always include variables that control for the performance measure used.

Second, focusing on the benefits accruing to the bottom of the income distribution ignores where in the remaining parts of the distribution the leaked benefits are going. For example, finding that a program is very ineffectively targeted at the bottom 20 percent is less worrying if the leaked benefits accrue mostly to people just above this income cut-off. This is partly why priority is given to the 40 percent measure of performance in constructing the performance index. It is also arguable that a focus of the bottom 40 percent coincides more closely with the objectives of most targeted programs. In any case, the fact that the results are extremely insensitive to the ordering is at least suggestive that where the cut-off point is drawn between 20 percent and 40 percent is somewhat inconsequential.

Third, recall that the data collated are only a sample of the hundreds of antipoverty interventions and that the performance indicator could be calculated only for two-thirds of this sample. These observations point to the possibility of sample selection bias—the possibility that certain characteristics of these programs (such as the fact that they were evaluated and documented) are themselves associated with the measures of targeting performance. A good example of this possibility relates to community targeting. The sample is only a fraction of the studies in Conning and Kevane (2001); it could well be that only successful interventions using community targeting have been well documented.

Fourth, some of the mistargeting observed here arises because households that were poor when the program admission decision was made were better off at the time of assessment or vice versa. This has implications for the design of targeted interventions. When there is considerable movement of households in and out of poverty and no mechanism for updating eligibility, methods that rely on static indicators of living standards (such as proxy-means tests) are likely to perform less well than methods that rely on self-selection.

Finally, the performance index focuses solely on the benefit side of the equation and ignores cost, which may be extremely important in selecting targeting methods or programs. For example, it is often argued that well-designed public works programs can be effective at concentrating benefits in the hands of the poor. But the high nontransfer costs associated with such programs (including nonwage costs and forgone income) substantially reduce their cost effectiveness.

Neglect of the cost side largely reflects data restrictions. Evidence on administrative costs was scant. Some cost data were available for 32 programs, but both cost and performance indicator data were available for only 20. Moreover, the cost data suffer from a severe lack of comparability. Most of the data for Latin America are

taken from Grosh (1994) and give administrative costs as a share of the program budget, based on official records. No attempt is made to determine how much of program benefits are siphoned off through corruption. Much of the cost data on South Asian programs are constructed from survey data on the value of benefits received by sample households. Based on the total program cost and appropriate grossing up, the total cost per dollar of benefit received is calculated. Corruption and theft appear to contribute more to total program expenses than legitimate administrative expenses, though little is said about these expenses. In any case, even when cost data are available, focusing on benefit incidence is extremely important in its own right.

It is worth reemphasizing that the objective of effectively targeting transfers, though always important, is often only one of the objectives of interventions. To the extent that there are tradeoffs between these other objectives and that of effective targeting, this needs to be taken into account in the overall evaluation of any program. It may be that these other objectives impinge more on program design and how the program is “sold” and delivered. Nonetheless, most policy analysts would accept that monitoring the targeting performance of programs dedicated mainly to poverty alleviation is always desirable, especially in developing economies, where poverty is high, budgets are tight, and other policy instruments (such as a comprehensive income tax system) are less developed, less sophisticated, and less progressive.

Regression Analysis

Although factors other than choice of method or program may be relatively large, this does not mean that these choices are unimportant. How important they are is shown by the results of a series of regressions that identify how performance varies systematically across these choices as well as across country characteristics (table 5). Targeting methods are themselves choices; they are not exogenous or predetermined. Consequently, it is incorrect to treat these results as causal relations. Rather, they are measures of partial correlation or association.

The first specification explores how country characteristics such as income, government accountability, and inequality are associated with (log) incidence. Income is measured as log gross domestic product (GDP) per capita (in purchasing power parity U.S. dollars) as of 1995.¹⁰ The hypothesis is that as a country becomes wealthier, it acquires the institutional capacity needed to design a well-targeted intervention.

Government accountability is based on work by Kaufmann, Kraay, and Zoido-Lobaton (1999), who define voice and accountability as a composite measure based on aspects of political processes, civil liberties, and political rights, capturing the

Table 5. Multivariate Analysis of Targeting Performance for Sample Programs in 1985–2003

Variable	Basic results				Dependent Variable in Levels	Uses median regression
	1	2	3	4	5	6
Log GDP per capita	0.120 (2.10)		0.181 (3.27)	0.150 (2.95)	0.201 (2.88)	0.133 (2.35)
Log voice	0.161 (2.06)			0.176 (2.88)	0.217 (2.52)	0.159 (1.65)
Log Gini coefficient	0.363 (1.75)			0.110 (0.39)	0.439 (1.17)	0.333 (1.14)
Means testing		0.284 (2.43)	0.285 (2.59)	0.300 (2.90)	0.397 (2.84)	0.306 (2.36)
Proxy-means testing		0.252 (1.43)	0.108 (0.74)	0.078 (0.45)	0.021 (0.10)	–0.054 (0.28)
Community assessment		0.198 (0.97)	0.119 (0.59)	–0.012 (0.09)	–0.017 (0.10)	–0.156 (0.59)
Geographic		0.260 (2.72)	0.336 (3.25)	0.341 (3.33)	0.381 (2.56)	0.370 (2.42)
Age, elderly		–0.055 (0.37)	–0.089 (0.68)	–0.117 (0.091)	–0.124 (0.075)	–0.145 (0.29)
Age, young		0.198 (1.98)	0.128 (1.34)	0.140 (1.53)	0.131 (0.94)	0.188 (1.89)
Other categorical		–0.132 (0.88)	0.028 (0.22)	0.137 (1.23)	0.061 (0.32)	0.088 (0.42)
Work		0.511 (3.66)	0.404 (3.74)	0.404 (3.73)	0.650 (2.83)	0.314 (2.22)
Community bidding		–0.018 (0.17)	–0.012 (0.11)	–0.029 (0.24)	–0.117 (0.69)	–0.149 (1.12)
F-statistic	3.46	3.56	5.41	7.20	5.89	
R ²	0.318	0.406	0.484	0.545	0.572	
Sample size	84	85	85	84	84	84

Note: The numbers in parentheses are the absolute values of *t*-statistics. All specifications contain controls, not reported, indicating whether the performance measure is based on proportion of benefits going to the poorest 20 percent, poorest 10 percent, the "poor," or the proportion of poor found in the population. Specifications 1–5 estimate standard errors using the methods proposed by Huber (1967) and White (1980). Specification 6 calculates standard errors using the bootstrap with 50 repetitions (uses median regression). Specifications 1–4 and 6 express the dependent variable in logs; specification 5 uses levels.

Source: Authors' compilation based on data search described in the text.

extent to which citizens participate in the selection of their governments and citizens and the media can hold governments accountable for their actions. They compiled subjective perceptions regarding the quality of governance in different countries, using such sources as polls of experts, commercial risk rating agencies, and cross-country

surveys. The analysis here used countries' rankings, which provide an easier way of interpreting the estimated coefficients. In the sample, Vietnam has the lowest percentile rank for voice (6) and Costa Rica the highest (88).¹¹

Country-specific Gini coefficients are included on the grounds that it might be easier to identify potential beneficiaries when income or consumption differences across individuals are greater. Also included (but not reported) are controls indicating whether the performance measure is based on the proportion of benefits going to the poorest 20 percent, the poorest 10 percent, the "poor" defined with reference to a poverty line, or the proportion of poor in the population. Doing so takes into account confounding effects arising from the use of different measures of poverty incidence in the studies on which this analysis is based. Standard errors are computed using the methods proposed by Huber (1967) and White (1980).¹²

The results shown in specification 1 show that as country income rises and as inequality rises, so does the targeting performance of antipoverty interventions (see table 5). Targeting is also better in countries where government accountability is better.

Specification 2 looks solely at the impact of choice of targeting method. Dummy variables are included for the nine targeting methods described: three kinds of individual assessment (means testing, proxy-means testing, community selection of individual beneficiaries), four kinds of categorical targeting (geographic, the elderly, the young, others), and two kinds of selection (work requirement, community bidding for projects). The omitted category is self-selection based on consumption. This was chosen as the base category for two reasons. Some argue that this form of targeting is a transition tool, to be used only until the capacity for more precise mechanisms—such as means testing—is developed.¹³ Others see self-selection based on the consumption of food as the preferred targeting mechanism and have expressed skepticism about the comparative ability of alternative targeting methods to reach the poor.¹⁴ Hence, an attractive feature of this specification is that the coefficients on these methods should be interpreted relative to self-selection based on consumption.

Specification 2 shows that means testing, geographic targeting, and self-selection based on a work requirement are all associated with an increased share of program resources going to the poorest 20 percent relative to self-selection based on consumption. Proxy-means testing and targeting the young are also associated with improved incidence, though these are measured with larger standard errors. Targeting the elderly, community assessment, other methods of categorical targeting, and selection based on community bidding are not associated with better incidence relative to the base category of self-targeting based on consumption.

Countries with greater capacity for program implementation may do better at directing benefits toward poorer members of the population either by choosing finer targeting methods or by implementing their choices more effectively. In such cases, the associations in specification 2 may be misleading, merely reflecting correlation

between unobserved implementation capacity and observed targeting methods. This possibility is explored in specifications 3 and 4. Controlling for country income level (specification 3) or income, voice, and inequality (specification 4) does not appear to eliminate the positive association—relative to self-selection based on consumption—between means testing, geographic targeting, and self-selection based on a work requirement and targeting performance. Targeting performance is better in countries with higher levels of income and where governments are held accountable for their actions. Specifically, a 10 percent increase in income is associated with a 1.8 percent increase in targeting performance. Raising the voice rank from 37 (Pakistan's rank) to 67 (India's rank) would be associated with about a 30 percent improvement in targeting performance.¹⁵ It is possible, however, that geographic targeting will also be more effective in countries with marked inequalities. Indeed when geographic targeting is dropped from the specification (but all other methods are retained), the parameter estimate for the log Gini coefficient is almost identical to that reported in specification 1.

Several additional specific checks were performed to investigate the robustness of this result. Specification 5 uses the same sample and regressors as specification 4, but the dependent variable is expressed in levels instead of logs. Basic results remain unchanged: means testing, geographic targeting, and targeting based on a work requirement improve targeting performance relative to the omitted category, self-selection based on consumption. There is no meaningful change in any of the other results. Specification 6 estimates median regressions, which express differences in performance in terms of differences in medians.¹⁶ This is an attractive check on robustness because the median is considerably less sensitive to outliers, an especially important consideration when working with small sample sizes. The results are broadly similar to those for specification 4—which uses an identical set of regressors, sample, and dependent variable—with the one exception being a markedly larger coefficient on the log Gini coefficient.¹⁷

The discussion has focused largely on the association between different targeting methods and targeting performance relative to self-selection based on consumption and conditioning on country characteristics. Also explored is the association between combinations of targeting methods and targeting performance (table 6). In addition to controls for income, voice, governance, inequality, and how the performance measure is constructed, the number of targeting methods used is added to specification 1. The results show that use of more methods is associated with improved targeting, each additional method improving performance by 15 percent. When a series of dummy variables is used in specification 2 to represent the number of targeting methods, the findings are similar. The sample size is too small to explore the association between specific groups of methods and targeting performance, but these results suggest that combining methods improves targeting.

Table 6. Association between Targeting Performance and Number of Methods Used for Sample Programs in 1985–2003

Variable	1	2
Number of methods used	0.165 (3.97)	
Used two methods		0.182 (1.66)
Used three methods		0.300 (2.88)
Used four or five methods		0.533 (3.11)
Log GDP per capita	0.141 (3.06)	0.141 (2.93)
Log voice	0.229 (3.49)	0.232 (3.07)
Log Gini	0.280 (1.35)	0.278 (1.31)
F statistic	6.42	5.08
R ²	0.426	0.419
Sample size	84	84

Note: Numbers in parentheses are absolute values of *t*-statistics. Specifications 1 and 2 contain controls, not reported, indicating whether the performance measure is based on proportion of benefits going to the poorest 20 percent, poorest 10 percent, the "poor," or the proportion of poor found in population. Specifications 1 and 2 estimate standard errors using the methods proposed by Huber (1967) and White (1980).

Source: Authors' compilation based on data search described in the text.

Conclusion

This article addresses the contested issue of the efficacy of targeting interventions in developing countries using a newly constructed database of 122 targeted anti-poverty interventions in 48 countries. These data are used to address three questions: What are the targeting outcomes? Are there systematic differences in targeting performance by targeting method and other factors? What are the implications of any such systematic differences for the design and implementation of targeted interventions?

The median value of the measure of targeting performance is 1.25, meaning that the median program transfers 25 percent more to poor individuals—persons and households in the bottom income groups—than would be the case with universal allocation. In this sense, targeting works. However, a staggering 21 of the 85 programs for which the performance measure can be built (25 percent) are regressive, with a performance index of less than 1. In these cases, random selection of beneficiaries would provide greater benefits to the poor. Some of this regressivity is driven by the inclusion of food subsidy interventions that use self-selection based on consumption as a targeting method. But even when these are dropped from the sample, 14 percent of targeted antipoverty interventions are still found to be regressive.

Countries with better capacity for program implementation, as measured by GDP per capita, do better at directing benefits toward poorer members of the population.

Countries where governments are more likely to be held accountable for their behavior—where “voice” is stronger—also appear to implement interventions with better targeting performance. Targeting is also better in countries where inequality is more pronounced and differences in economic well-being are presumably easier to identify.

Mindful of caveats, interventions that use means testing, geographic targeting, and self-selection based on a work requirement are all associated with an increased share of benefits going to the poorest 40 percent. Proxy-means testing, community-based selection of individuals, and demographic targeting to children show good results on average, but with wide variation. Self-selection based on consumption, demographic targeting to the elderly, and community bidding show limited potential for good targeting.

That said, there is considerable variation in targeting performance when experiences with specific program types and targeting methods are examined. This partly reflects heterogeneity across the interventions assessed because some have multiple objectives whereas others are focused solely on poverty reduction. But when targeting by specific method or program type is considered, there is considerable intra-group variability. Thus, it is not surprising that although community assessment generally performs no better than self-targeting based on consumption, Alderman's (2002) study of community targeting in Albania describes a highly successful example of this form of targeting. Similarly, Case and Deaton (1998) and Duflo (2000) show that in South Africa targeting the elderly is an effective method for reaching poor children, even though targeting the elderly generally performs relatively poorly when compared with other methods for reaching poor people.

Thus, although the patterns observed are instructive, they should not be interpreted as a lexicographic ranking of methods. Differences in individual country characteristics and implementation are also important determinants of outcomes and must be considered carefully in making targeting decisions. For example, in countries where illiteracy is high, it makes little sense to require potential beneficiaries to fill out forms as part of a means test. Such requirements discourage otherwise eligible beneficiaries from applying, as Munro (2003) illustrates for Zimbabwe. This suggests that work on targeting should extend beyond simple quantitative comparisons of methods to consider more detailed and often qualitative issues of comparisons within methods. How does and should implementation differ in different settings, and how can constraints of political economy, poor information, or low administrative capacity best be accommodated or reduced? In a companion article, Coady, Grosh, and Hoddinott (forthcoming) provide a more detailed discussion of the merits, limitations, and options for implementing individual targeting methods in an attempt to move in this direction.

Notes

David Coady is research fellow in the Food Consumption and Nutrition Division at the International Food Policy Research Institute; his e-mail address is d.coady@cgiar.org. Margaret Grosh is lead economist in the Social Protection Anchor Unit at the World Bank; her e-mail address is mgrosh@worldbank.org. John Hoddinott is senior research fellow in the Food Consumption and Nutrition Division at the International Food Policy Research Institute; his e-mail address is j.hoddinott@cgiar.org. The authors thank Akhter Ahmed, Harold Alderman, John Blomquist, Stephen Devereux, Lant Pritchett, Kalanidhi Subbarao, and Salman Zaidi for comments and reference materials; two anonymous referees for helpful comments; Yisgedu Amde and Sanjukta Mukherjee for research assistance; and Manorama Rani for document processing. They also thank participants in seminars at the International Food Policy Research Institute and the World Bank, a conference on Chronic Poverty and Development Policy at the University of Manchester in April 2003, and the North Eastern Universities Development Consortium Conference 2004 at Yale University for useful remarks. The study was commissioned by the World Bank's Social Protection Anchor Unit for the Safety Nets Primer series. Other papers are available online at www.worldbank.org/safetynets.

1. Van de Walle (1998) contains a useful discussion of these approaches.
2. General discussions of the principles underlying narrow targeting are found in Atkinson (1995), Besley and Kanbur (1993), Grosh (1994), van de Walle (1998), and Coady, Grosh, and Hoddinott (forthcoming).
3. This focus is conditioned by three factors. First, if targeting is largely ineffective, the answers to the remaining questions are moot. Second, there are simply not enough studies with cost data, and the cost data that are available suffer severely from lack of comparability. Third, assessment of impact requires careful attention to the counterfactual or what beneficiaries would have done in the absence of these interventions. Few studies of transfer programs in developing countries do that with any care; exceptions are Datt and Ravallion (1994), Ravallion and Datt (1995), Jalan and Ravallion (1999), and Skoufias (2001).
4. The database is available in the form of an annotated bibliography (Coady, Grosh, and Hoddinott 2003) that includes details on the study itself (title, authors, reference details, year of publication, study objective), background information on the intervention (program name, year implemented, program description, type of benefit, program coverage and budget, transfer levels), targeting method (what criteria were used to determine eligibility, targeting mechanism), how the intervention operated, targeting performance (who benefited), and descriptions of impact on welfare and costs of targeting.
5. Also, the microcredit literature is vast. Attempting to construct a database comparable to that developed here is a major undertaking beyond the scope of this analysis.
6. Strictly speaking, all programs are self-targeted to some degree, because there are always some actions (and therefore costs) required of beneficiaries to register for and collect a benefit.
7. Social funds also use other mechanisms, such as geographical targeting. Differences in access to information or capacity for demanding social funds also account for differential access to these interventions.
8. Alderman and Lindert (1998) provide a recent review of the potential and limitations of self-targeted food subsidies.
9. How these rankings are calculated varies by study. In general, they are based on nationally representative household survey data that estimate the share of transfers that accrue to individuals in the bottom deciles of the consumption distribution. Such estimates do not take into account changes in private transfer behavior or labor supply. Coady, Grosh, and Hoddinott (2003) provide a description of these calculations by intervention as well as references to the original source material.
10. Income and inequality data are from the World Bank's World Development Indicators database.
11. Preliminary work included a measure of government effectiveness also drawn from Kaufmann, Kraay, and Zoido-Lobaton (1999). However, it is never statistically significant, quite possibly because it is highly correlated with log per capita income.
12. Following the suggestion of an anonymous referee, we explored whether the studies from which data were drawn for this analysis had been published in a refereed journal, a book, or was unpublished work undertaken by World Bank staff, International Food Policy Research Institute (IFPRI)

staff, or individuals based in other institutions. These controls can be thought of in two ways. Arguably, work published in journals (and possibly books) has been more rigorously reviewed, so those studies will be less dogged by measurement error. On the other hand, there may be publishing biases in that only studies with “significant results” are submitted and accepted by journals. Of the 85 estimates of targeting performance used here, 17 come from refereed journals, 17 from books, 37 from unpublished work by World Bank staff, 8 from unpublished work by IFPRI staff, and 6 from unpublished work by researchers based at other institutions. When dummy variables for type of publication are included as additional controls, they are not found to be jointly statistically significant and hence are not reported here. But coefficients on dummy variables for unpublished studies by World Bank staff or by individuals based in other institutions tend to be negative, consistent with the publishing bias hypothesis.

13. See, for example, Pinstrup-Andersen (1988) and Alderman and Lindert (1998).

14. Such implicit concern is found, for example, in Cornia and Stewart (1995).

15. Kaufmann, Kraay, and Zoido-Lobaton (1999) caution that these composite measures are likely to be measured with error. As such, they are likely to provide lower-bound estimates of the impact of these characteristics.

16. More precisely, a quartile regression was estimated centered at the median with standard errors obtained via bootstrap resampling with 50 repetitions to correct for heteroscedasticity. Increasing the number of repetitions does not appreciably alter the standard errors.

17. As a further specification check, specification 4 was reestimated, but the sample was restricted in three ways: including only studies reporting the share of benefits accruing to the bottom 40 percent, including only studies reporting the share of benefits accruing to the poorest 40 percent or the poorest 20 percent, and including only studies that report the share of benefits accruing to the poorest 40 percent, the poorest 20 percent, or the poorest 10 percent. The results, available on request, are comparable to those reported in specification 4.

References

- Alderman, Harold. 2002. “Do Local Officials Know Something We Don’t? Decentralization of Targeted Transfers in Albania.” *Journal of Public Economics* 83(3):375–404.
- Alderman, Harold, and Kathy Lindert. 1998. “The Potential and Limitations of Self-targeted Food Subsidies.” *World Bank Research Observer* 13(2):213–29.
- Atkinson, Anthony. 1995. “On Targeting Social Security: Theory and Western Experience with Family Benefits.” In Dominique van de Walle and Kimberly Nead, eds., *Public Spending and the Poor*. Baltimore, Md.: Johns Hopkins University Press.
- Besley, Timothy, and Ravi Kanbur. 1993. “Principles of Targeting.” In M. Lipton and J. van der Gaag, eds., *Including the Poor*. Washington D.C.: World Bank.
- Bigman, David, and Hippolite Fofack. 2000. “Geographic Targeting for Poverty Alleviation: An Introduction to the Special Issue.” *World Bank Economic Review* 14(1):129–46.
- Braithwaite, Jeanine, Christiaan Grootaert, and Branko Milanovic. 2000. *Poverty and Social Assistance in Transition Countries*. New York: St. Martin’s Press.
- Case, Anne, and Angus Deaton. 1998. “Large Cash Transfers to the Elderly in South Africa.” *Economic Journal* 108(4):1330–61.
- Coady, David, and Emmanuel Skoufias. 2001. “On the Targeting and Redistributive Efficiencies of Alternative Transfer Programs.” FCND Discussion Paper 100. International Food Policy Research Institute, Washington, D.C.
- Coady, David, Margaret Grosh, and John Hoddinott. 2003. “Targeted Anti-poverty Interventions: A Selected Annotated Bibliography.” International Food Policy Research Institute, Washington, D.C.

- . Forthcoming. *The Targeting of Transfers in Developing Countries: Review of Experience and Lessons*. Regional and Sectoral Studies Series. Washington, D.C.: World Bank.
- Conning, Jonathan, and Michael Kevane. 2001. "Community Based Targeting Mechanisms for Social Safety Nets." Social Protection Discussion Paper 102. World Bank, Washington, D.C.
- Cornia, Giovanni, and Frances Stewart. 1995. "Two Errors of Targeting." In Dominique van de Walle and Kimberly Nead, eds., *Public Spending and the Poor*. Baltimore, Md.: Johns Hopkins University Press.
- Datt, Gaurav, and Martin Ravallion. 1994. "Transfer Benefits from Public-Works Employment: Evidence for Rural India." *Economic Journal* 104(427):1346–69.
- Dufo, Esther, 2000. "Child Health and Household Resources in South Africa: Evidence from the Old Age Pension Program." *American Economic Review* 90(2):393–98.
- Grosh, Margaret. 1994. *Administering Targeted Social Programs in Latin America: From Platitudes to Practice*. Washington D.C.: World Bank.
- Huber, Peter. 1967. "The Behavior of Maximum Likelihood Estimates under Non-standard Conditions." In Lucien M. LeCam and Jerzy Neyman, eds., *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*. Vol. 1. Berkeley, Calif.: University of California Press.
- Jalan, Jyotsna, and Martin Ravallion. 1999. "Income Gains to the Poor from Workfare: Estimates for Argentina's Trabajar Program." Policy Research Working Paper 2149. World Bank, Development Research Group, Poverty and Human Resources, Washington, D.C.
- Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobaton. 1999. "Aggregating Governance Indicators." Policy Research Working Paper 2195. World Bank, Development Research Group, Macroeconomics and Growth, and World Bank Institute, Governance, Regulation, and Finance, Washington, D.C.
- Munro, Lauchlan. 2003. "A Social Safety Net for the Chronically Poor? Zimbabwe's Public Assistance in the 1990s." Fordham University, Department of Policy, Planning, and Development, New York.
- Pinstrup-Andersen, Per, ed. 1988. *Food Subsidies in Developing Countries: Costs, Benefits and Policy Options*. Baltimore, Md.: Johns Hopkins University Press.
- Ravallion, Martin. 1993. "Poverty Alleviation through Regional Targeting: A Case Study for Indonesia." In Karla Hoff, Avishay Braverman, and Joseph Stiglitz, eds., *The Economics of Rural Organization: Theory Practice and Policy*. New York: Oxford University Press.
- Ravallion, Martin, and Calvin Chao. 1989. "Targeted Policies for Poverty Alleviation under Imperfect Information: Algorithms and Applications." *Journal of Policy Modeling* 11(2):213–24.
- Ravallion, Martin, and Gaurav Datt. 1995. "Is Targeting through a Work Requirement Efficient?" In Dominique van de Walle and Kimberly Nead, eds., *Public Spending and the Poor*. Baltimore, Md.: Johns Hopkins University Press.
- Rawlings, Laura, Lynne Sherburne-Benz, and Julie Van Domelen. 2001. "Letting Communities Take the Lead: A Cross-Country Evaluation of Social Fund Performance." World Bank, Washington, D.C.
- Skoufias, Emmanuel. 2001. "PROGRESA and Its Impacts on the Human Capital and Welfare of Households in Rural Mexico: A Synthesis of the Results of an Evaluation by IFPRI." International Food Policy Research Institute, Washington, D.C.
- van de Walle, Dominique. 1998. "Targeting Revisited." *World Bank Research Observer* 13(2):231–48.
- White, Halbert. 1980. "A Heteroscedasticity-Consistent Covariance Matrix and a Direct Test for Heteroscedasticity." *Econometrica* 48(4):817–38.
- World Bank. 1990. *World Development Report 1990: Poverty*. New York: Oxford University Press.
- . 1997. *World Development Report 1997: The State in a Changing World*. New York: Oxford University Press.
- . 2001. *World Development Report 2000/2001: Attacking Poverty*. New York: Oxford University Press.

