Targeting outcomes, redux

David Coady
Margaret Grosh
John Hoddinott

November 2002

Acknowledgements
We thank Akhter Ahmed, Harold Alderman, John Blomquist, Stephen Devereux, Lant Pritchett, Kalanidhi Subbarao and Salman Zaidi for useful comments and supplying us with reference materials ,Yisgedu Amde and Sanjukta Mukherjee for helpful research assistance and Manorama Rani for document processing. We also thank seminar participants at IFPRI and the World Bank for their very useful remarks. The paper was commissioned by the Social Protection Anchor unit for the Safety Nets Primer series. The findings, interpretations and conclusions expressed in this publication are those of the authors and should not be attributed in any manner to the World Bank, its affiliated organizations, or to the members of the Board of Directors or the countries they represent.

Address for correspondence
Coady, Hoddinott: FCND, International Food Policy Research Institute, 2033 K Street NW Washington DC 20006, USA. em: D.Coady@cgiar.org; J.Hoddinott@cgiar.org
Grosh: World Bank, 1818 H Street, NW Washington DC 20433, USA. em: MGrosh@worldbank.org

Abstract

This paper addresses the contested issue of the efficacy of targeting interventions in developing countries using a newly constructed comprehensive database of 111 targeted anti-poverty interventions in 47 countries. While the median program transfers 25 percent more to the target group than would be the case with a universal allocation, more than a quarter of targeted programs are regressive. Countries with higher income or governance measures and countries with better measures for voice do better at directing benefits towards poorer members of the population. 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. Self-selection based on consumption, demographic targeting to the elderly, and community bidding show limited potential for good targeting. Proxy means testing, community-based selection of individuals and demographic targeting to children show good results on average, but with considerable variation. Overall, there is considerable variation in targeting performance when we examine experiences with specific program types and specific targeting methods. Indeed a Theil decomposition of the variation in outcome shows that differences between targeting methods account for only 20 percent of overall variation, the remainder is due to differences found within categories. So while these general patterns are instructive, differences in implementation are also quite important determinants of outcomes.

1. Introduction

Over the last two decades there has been an emerging consensus that while economic growth is a necessary condition for alleviating poverty within an acceptable timeframe, in isolation it is not sufficient (World Bank, 1990, 1997, 2000). First, the asset base of poor households needs to be built up so that they can participate in the growth process. Second, growth needs to be more intensive in the assets held by the poor and the sectors in which they predominate. Third, because it takes time for the benefits from such a strategy to accrue, short-term public transfers are required to protect and raise the consumption of the poorest households.

Implementation of this agenda for reducing poverty requires methods for reaching the poor. This can be accomplished by 'broad targeting' in the form of spending on items that reach a wide swath of society including the poor (for example, universal primary education, an extensive network of basic health care) or by 'narrow targeting' where methods that identify the poor more specifically are used to confer benefits only to them (for example, transfer programs) (van de Walle, 1998). The case for the latter form of targeting arises from the existence of a budget constraint (Besley and Kanbur, 1993).[1] The overall poverty impact of a program depends both 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 non-poor households is a lower impact in terms of poverty reduction, reflecting less coverage of poor households and/or lower benefit levels. By targeting transfers to poor households, one can increase the amount transferred to them.

In addition to the debate surrounding the appropriate balance between broadly and narrowly targeted interventions, there are sharply divergent views as to how much the latter actually benefit the poor. Divergent views on the efficacy of this approach are based on differing assessments of three questions: i) Whether better targeting outcomes are likely to be achieved; ii) Whether such methods are cost effective; and iii) Whether the living standards of the poor are improved by such targeted interventions. This paper addresses the first question.[2] While it would seem that there exists a fairly extensive literature on this topic, it is largely dominated by descriptions of individual, sometime idiosyncratic, programs. Even comparative analyses tend to cover either a single region (eg. Grosh, 1994, for Latin America and the Caribbean; Braithwaite, Grootaert and Milanovic, 2000 for Eastern Europe and Central Asia), or method (Bigman and Fofack,

[1] General discussions of the principles underlying narrow targeting are also found in Atkinson (1995), Grosh (1994), van de Walle (1998) and Coady, Grosh and Hoddinott (2002a).

[2] We stress that this focus does not arise because we consider the second and third questions to be unimportant. Rather, our focus in whether targeted interventions reach the poor is conditioned by three factors. First, if targeting is largely ineffective, the answers to these remaining questions are moot. Second, there are simply not enough studies with cost data. As we discuss in the paper, fewer than 20 per cent of the interventions in our database report information on both targeting performance and the cost of targeting. Moreover, the cost data suffer severely from lack of comparability. Third, assessment of impact requires careful attention to the counterfactual, what beneficiaries would have done in the absence of these interventions. Few studies do so with any care, exceptions being Datt and Ravallion (1994), Ravallion and Datt (1995) and Jalan and Ravallion (1999).

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.

We rectify this weakness by drawing on a newly constructed database of 111 targeted anti-poverty interventions drawn from 47 countries in Latin America and the Caribbean, Eastern Europe and the Former Soviet Union, the Middle East and North Africa, Sub-Saharan Africa, and South and East Asia. We use these data to address three questions: 1) What targeting outcomes are observed? 2) Are there systematic differences in targeting performance by targeting methods and other factors? 3) What are the implications for such systematic differences for the design and implementation of targeted interventions?

We find that the median targeted program is progressive in that it transfers 25 percent more to the target group than would be the case with a universal (or random) allocation. However, for a staggering quarter of the programs outcomes are regressive. Countries with higher income or measures of governance, which we take to imply better capacity for program implementation do better at directing benefits towards poorer members of the population as do countries where governments are more likely to be held accountable for their behaviour, as suggested by better measures of voice. Targeting is also better in countries where inequality is more pronounced.

Subject to a number of caveats, we find that relative to self-selection based on consumption (used for example in universal food subsidy schemes) 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. Self-selection based on consumption, demographic targeting to the elderly, and community bidding show limited potential for good targeting. Proxy means testing, community-based selection of individuals and demographic targeting to children show good results on average, but with wide variation. That said, we emphasize that there is considerable variation in targeting performance when we examine experiences with specific program types and specific targeting methods. Indeed a Theil decomposition of the variation in outcome shows that differences between targeting methods account for only 20 percent of overall variation, the remainder is due to differences found within categories. How well a program implements a chosen targeting method is as important as which method is chosen.

2. Data Construction and Description

2.1 Database Construction

As noted above, while there is a fairly rich literature on targeted programs, much of it either documents single programs or compares outcomes within a single region, method or class of intervention. Accordingly, the first step in our analysis was an extensive literature review and the construction of a database of targeted anti-poverty interventions.[3] To our knowledge, this work represents the most extensive attempt to construct such a database.

Our criteria for inclusion in this database were the following: i) The intervention had to be situated in a low or middle-income country; ii) A principal objective of the intervention is poverty reduction defined in terms of income or consumption; iii) Documentation on the intervention contains information on the targeting method used, its implementation and something about outcomes; and iv) The intervention is relatively recent (generally from 1985-2002). Included in our data are cash transfers (including welfare and social assistance payments, child benefits and non-contributory pensions), near-cash transfers (such as quantity rationed subsidized food rations and food stamps), food transfers, universal food subsidies, non-food subsidies, public works, and social funds.

Two observations should be made on these criteria for inclusion. First, a number of interventions that are included have objectives that include, but are not limited to, direct poverty reduction. Social funds are a good example. While short term poverty reduction can be an important component of these interventions, so too 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 "anti-poverty" interventions means that one must be cautious in interpreting comparisons across types of interventions.

Second, focusing the review in this way necessarily means excluding a number of interventions that may, in some cases, be targeted, and may have some impact in terms of poverty reduction. Thus, excluded are: 'occupationally based transfer schemes' such as formal sector unemployment insurance or occupational old age or disability pensions (here, the principal mechanism that determines eligibility and benefit levels are employment and contributions history rather than current poverty status); credit and micro-credit schemes (although these are often targeted, they are motivated, in large part, by credit market failures); supplementary feeding programs (mainly because our foray into the vast literature on this type of intervention did not yield studies that satisfied the criteria described above); and most short-term emergency aid (because although this has a clear poverty focus, and is often targeted by need, the time scale on which it operates typically precludes an assessment of the distribution of the benefits).

Most studies of targeting – especially those outside of Latin America and the Caribbean - do not appear in peer-reviewed journals. Consequently, we undertook searches of the "gray" literature using Web search engines found at the World Bank, ELDIS and IFPRI web sites using the following key words: safety nets, targeting, social funds, pensions,

[3] This is available in the form of an annotated bibliography, Coady, Grosh and Hoddinott (2002b). For each program we obtained 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.

public works, subsidies. Additional cases were found via canvassing colleagues about work that had not yet been catalogued in these places. Searches were also undertaken in the following academic journals for the years 1990-2002: *Economic Development and Cultural Change*, the *Journal of Development Economics*, the *Journal of Development Studies*, the *Journal of Public Economics*, the *World Bank Economic Review*, the *World Bank Research Observer* and *World Development,* and *Economic and Political Weekly* for 1998-2002. Additional cases were found through reviews of existing compilations such as Grosh (1994) and Braithwaite, Grootaert and Milanovic (2000).

Given the nature of such a search, it is important to remember that our sample of interventions is not necessarily reflective of the distribution of programs that exist in the world, but rather of those that have some measurement of targeting outcomes that has been written up in the catalogued English language literature that we have been able to access. Programs are 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, it is a program that by virtue of methods or setting is deemed attractive by analysts and editors. For example, we suspect that programs using community based methods and agents are under-represented. These are often only locally funded and the methods chosen when there are poor data and low administrative capacity, features which all reduce the likelihood of an evaluation being done and finding its way into the international literature. For similar reasons, it is likely that we under-represent the literature on public works in sub-Saharan Africa. Proxy-means tests are, on the other hand, well represented, with a large share of all such programs in the world showing up in this sample.

2.2 Data Description

Based on the criteria described above, we were able to collect information for 111 interventions drawn from 47 countries. Table 1 provides a description of the distribution of these interventions across regions, income groups and intervention types. We can see that this sample of interventions provides a fairly broad regional coverage. Although cash transfer programs account for a large proportion (38%) of the interventions, the other intervention types are well represented. In some regions, a particular intervention type dominates: cash transfers in Eastern Europe, the Former Soviet Union and Central Asia (ECA), universal food subsidies in the Middle East and North Africa (MENA), and near-cash transfers in South Asia (SEA). By contrast, there is a wider mix of reported interventions in other regions. Most of the cash transfer programs occur in Latin America and the Caribbean (LAC) and ECA, most of the near-cash transfer programs occur in South Asia, most of the universal food subsidies occur in MENA, and most of the social funds occur in LAC. Dividing the sample by per capita GDP levels, we find that cash transfer programs are more likely to be found in less poor countries and near-cash transfers in the poorest countries.

Table 2 provides information on the distribution of interventions and their targeting methods. We distinguish between three broad forms of targeting: individual/ household assessment, categorical, and self-selection and various sub-categories within each of these.

Individual/ household assessment: is a method under which eligibility is directly assessed on an individual basis. In a verified means test, (nearly) complete information is obtained on a household's income and/or wealth and compared to other sources of information such as pay stubs, or income and property tax records. This requires the existence of such verifiable records in the target population, as well as the administrative capacity to process this information, and to continually update it, in a timely fashion. Absent the capacity for a verified means test, other individual assessment mechanisms are used. For example, simple means tests, with no independent verification of income, are not uncommon. A visit to the household may help to verify in a qualitative way that visible standards of living (which reflect income or wealth) are more or less consistent with the figures reported. Proxy-means tests involve generating a score for applicants based on fairly easy to observe characteristics of the household such as the location and quality of the 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 so that a group of community members or a community leader whose principal functions in the community are not related to the transfer program will decide who in the community should benefit and who should not.

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 to be easy to observe, hard to falsely manipulate, and correlated with poverty. Age, gender, ethnicity, land ownership, household demographic composition or location, are common examples. Geographic targeting is often used and often in tandem with other methods.

Self-selection: With some interventions, although eligibility is universal the design intentionally involves dimensions that are thought to encourage the poorest to use the program and the non-poor not to do so.[4] This is accomplished by recognizing differences in the private participation costs between poor and non-poor households. For example, this may involve: (a) 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 jobs; (b) the restriction of transfers to take place at certain times with a requirement to queue; or (c) or the location of points of service delivery in areas where the poor are highly concentrated so that the non-poor have higher (private and social) costs of access. An alternative form of self-selection is found in social fund-type interventions where communities apply for program funds. Here, selection uses

[4] Note because there are always some actions (and therefore costs) required of beneficiaries in order to register for and collect a benefit, strictly speaking all programs are self-targeted to some degree.

differences in the private participation costs between poor and non-poor communities as a way of targeting benefits.[5]

Note that universal food subsidies (with or without quantity rationing) can be viewed as a form of self-selection since they are universally available and households receive benefits by deciding to consume the commodity. In practice households can often determine not just whether or not to participate, but also the intensity of their participation. The more income elastic are expenditures on these items the more effective is the targeting. For example, food transfers often involve commodities with "inferior" characteristics (e.g. low quality wheat or rice) and households often substitute away from such expenditures as incomes increase.[6]

Table 2 uses this broad taxonomy of targeting methods but also specifies the principal approaches taken within the three broad categories of individual assessment, categorical targeting and self-selection. The first thing to notice is that interventions use a combination of targeting methods; in all we have 226 occurrences of different targeting methods, so that the interventions in our sample use just over two different targeting methods on average. Just 37 interventions use a single targeting method, while 43 use two methods, 21 use three methods, and 10 use four methods.

There are some marked differences by region. Most of the interventions using means- and proxy-means testing are concentrated in ECA and LAC. A legacy of the central planning era in ECA has been an extensive administrative system that is suited to the individual assessment of individual circumstances 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 in this region is based either on some form of individual assessment or individual characteristic such as age. Reliance on food subsidies explains why self-targeting based on consumption patterns is the dominant targeting method in MENA. SEA is notable for its extensive use of geographic targeting as well as a relatively high reliance on self-selection based on work or consumption. LAC countries also use geographic targeting extensively, but this is more often accompanied by either direct individual assessment (i.e. means or proxy means testing) or by targeting children. The small number of documented programs for sub-Saharan Africa (SSA) and East Asia and the Pacific (EAP) show more mixed patterns.

There are also broad differences across income levels. Generally, poorer countries tend to rely more on self-selection methods and categorical targeting whereas forms of individual assessment 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.

---

[5] Social funds also use other mechanisms such as geographical targeting. Differences in access to information or capacity for 'demanding' social funds also accounts for differential access to these interventions.

[6] Alderman and Lindert (1998) provide a recent review of the potential and limitations of self-targeted food subsidies.

Although certain program types are synonymous with certain targeting methods, most use a combination of methods, presumably because there is synergy from the perspective of targeting efficiency. Public works programs typically use a combination of geographic targeting and self-selection based on low wages and a work requirement. But, in practice, public works also often require additional rationing of employment using categorical targeting if demand exceeds supply at the wage paid. 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).

3. Assessing targeting effectiveness

In this section we describe methodologies used to evaluate the targeting efficiency of anti-poverty interventions. We outline the methodology used in this paper to compare targeting performance across interventions, identifying some important caveats that must be kept in mind when interpreting this indicator. We also provide a brief description of targeting outcomes in terms of this indicator of targeting performance.

3.1 Methods

A common approach to evaluate the targeting performance of alternative transfer instruments is to compare leakage and undercoverage rates. *Leakage* is the proportion of those who are reached by the program (i.e. are "in" denoted by *i*, as opposed to "out of", denoted by *o*, the program) who are classified as non-poor (errors of inclusion) or:

$$L \; ? \; \frac{N_{np,i}}{N_i}$$

where $N_{np,i}$ is the number of non-poor households in the program and N_i is the total number of households in the program.

Under-coverage is the proportion of poor households who are not included in the program (errors of exclusion), or:

$$U \; ? \; \frac{N_{p,o}}{N_p}$$

where $N_{p,o}$ is the number of poor households who are left out of the program and N_p is the total number of poor households.

There are two obvious criticisms of this approach (Coady and Skoufias, 2001). Firstly, it ignores the seriousness of the targeting errors in so far as: (i) it does not differentiate between the erroneous inclusion of non-poor households lying just above the poverty line and those lying well above the line, and (ii) it does not differentiate between the erroneous exclusion of poor households just below the poverty line and those well below the poverty line. In both cases, the different errors are identically treated. Secondly, it focuses only on who gets the transfers and not on how much households get (i.e. the size of the transfer budget). Thirdly, when comparing across programs it is often the case that those that do well on under-coverage simultaneously score badly on leakage. For example, so-called universal programs would be expected to score relatively well on under-coverage but badly on leakage, but the leakage/undercoverage approach does not address the issue of trade-off. Much of the problem with this approach therefore lies in the fact that the relative social valuation of income transfers to different households (e.g. moderately versus extremely poor) is not made explicit, although it is obvious that all the poor are treated similarly and all the non-poor are also treated similarly even if the issue of their relative weights is ignored.

Another commonly used approach to evaluating the effectiveness of targeting can be viewed as an attempt to incorporate the size of transfers and the budget explicitly into the analysis as well as how transfer levels are differentiated across households in different parts of the income distribution. Rather than asking how effective the program is at identifying the poor, it asks how effective it is at reducing poverty. It proceeds by comparing the relative impacts of the alternative instruments on the extent of poverty subject to a fixed common budget or, equivalently, the minimum cost of achieving a given reduction in poverty across instruments (Ravallion and Chao, 1989; Ravallion, 1993).

An alternative performance index is the *distributional characteristic* more commonly used in the literature on commodity taxation (Newbery and Stern, 1987; Ahmad and Stern, 1991; Coady and Skoufias, 2001). This is defined as:

$$? \; ? \; \frac{\sum_h ?^h T^h}{\sum_h T^h} \; ? \; \sum_h ?^h ?^h$$

where $ß^h$ is the social valuation of income transferred to household *h* (or its "welfare weight), T^h is the level of the transfer to the household, and $?^h$ is each household's share of the total program budget. The attraction of this index is that welfare weights are made explicit. For example, if poor households are given a welfare weight of unity and non-poor households a weight of zero, and we further assume that all beneficiary households receive the same level of transfer, then this index collapses to (1-L), the proportion of households receiving transfers that are classified as poor. If, in addition, we know the level of benefits received by beneficiaries, then it collapses to the share of the program budget received by poor households. Where the "poor" are defined as households falling within the bottom deciles (e.g. 20 percent or 40 percent) of the national income distribution, similar indices can be calculated. Generally, all that is required to calculate the distributional characteristic is mean incomes by decile and decile shares in transfers. The administrative cost side of the program can also be easily incorporated by including this cost in the denominator along with total transfers.

3.2 Our measure of targeting effectiveness

In order to compare the performance of the different targeting methods used in the range of programs considered in our analysis, we need a comparable performance indicator for each program. As is always the case is such "meta" analyses, the definitions, methods and presentations in the original studies vary in ways that make it difficult to assemble such 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/non-poor classification that differs by country. Other studies report none of these measures, but use other less common ones. And, of course, the measure of welfare used is not always strictly comparable from study to study. Thus we are faced with how best to compare targeting performance outcomes using data that are not strictly comparable.

Most studies catalogued in our database provide information on at least one of the following indices:
?? The proportion of total transfers received by households falling within the bottom 40, 20 or 10 percent of the national income distribution.
?? The proportion of beneficiaries falling within the bottom 40, 20, or 10 percent of the national income distribution.
?? The proportion of total transfers or beneficiaries going to "poor" households, where the poor are defined in terms of some specified part of the welfare distribution (e.g. falling in the bottom 35 percent of the income distribution).

As indicated above, ideally we would like to know the proportion of total transfers received by households falling within different centiles (40th, 20th, 10th and so on) of the national income distribution. This is a better measure that the proportion of beneficiaries by centile because in the case of the latter, we do not necessarily know anything about variations in the levels of transfers. These two measures – proportions of total transfers and proportions of beneficiaries – are only equivalent when transfer levels are uniform across beneficiaries.

Given that no single common measure of targeting performance is available, we have constructed a measure based on a comparison of actual performance to a common reference outcome, namely, 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. One can think of neutral targeting as arising either from the random allocation of benefits across the population or a universal intervention in which all individuals received identical benefits. The indicator used in our analysis is constructed by dividing the actual outcome by the appropriate neutral outcome. For example, if the bottom 40 percent of the income distribution receive 60 percent of the benefits then our indicator of performance is calculated as (60/40)=1.5, thus a higher value is associated with better targeting performance. A value of 1.5 means that targeting has led to the target group (here those 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 less than one for regressive targeting, with unity denoting neutral targeting.

The performance indicator used in the analysis below is based on a lexicographic selection process among the available incidence indicators based on the different "target groups". Where it is available, we base performance on the proportion of benefits accruing to the bottom two quintiles. Where this is not available, we base it on the proportion of benefits accruing to the bottom quintile, then benefits to the bottom decile and lastly, the share of program benefits received by individuals deemed to be below a poverty line. We can calculate such a performance indicator for 77 programs.

3.3 Descriptive Results

Table 3 lists all programs for which we can construct our performance indicator from best to worst. There is enormous variation in targeting performance, ranging from 4, for the Trabajar public works program in Argentina to 0.28, for VAT exemptions on fresh milk in South Africa. The median value is 1.25, so that the "typical" program transfers 25 percent more to the target group than would be the case with a universal (or random) allocation. However, a staggering 21 of the 77 programs – more than a quarter -- are regressive, with a performance index less than one. In these cases, a random selection of beneficiaries would actually provide greater benefits to the poor.

It is instructive to focus on the worst and best ten programs. The worst ten have a median score of only 0.64, ranging from 0.28-0.85, and are mainly from SSA and the MENA, with three from South Africa's VAT exemption program. Seven out of the ten are food subsidy programs, and two of the remaining three programs involve cash transfers. In fact, median performance rises to 1.35 if interventions using self-selection based on consumption are withdrawn from the sample. Doing so also reduces the proportion of regressive interventions to 16 per cent. It is also noticeable that only one of the poorly performing programs use either means or proxy-means targeting methods, none of them are geographically targeted, and they come from across the income spectrum. The top ten programs have a median score of 2.1, range from 1.95-4.0, and are from either LAC or ECA. Seven out of the ten involve cash transfers. Nine out of the ten make use of means, proxy-means, or geographic targeting, and seven out of the ten are in less-poor countries.

The fact that cash-transfers feature in both the best and worst ten programs highlights the possibility that variations in targeting performance may reflect poor implementation rather than poor potential for such program *per se.* It is, however, noticeable that whereas public works are all in the top half of the performance table, social funds are nearly all in the bottom half. This is consistent with there being a trade-off between the objective of reducing 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 dominance of less poor countries among the top half of the table suggests that characteristics correlated with income such as administrative capacity are important determinants of targeting performance.

Table 4 develops this idea further by providing summary statistics on targeting performance – sample size, median, interquartile range (*iqr*) and the *iqr* as a percentage of the median - by targeting method. First impressions suggest that Table 4 yields a clear hierarchy in terms of targeting performance. Interventions using forms of individual assessment have better incidence than those relying on forms of categorical targeting which in turn out-perform interventions that use self-selection much as one would expect. A closer inspection, however, reveals that such impressions are too general to be very useful. First, there is much heterogeneity within these broad methods of targeting. Most notably, the category of self-selection includes interventions utilizing a work requirement that have the highest median performance and self-selection based on consumption, which has the lowest median. Second, three specific methods – categorical targeting to the elderly, self-selection 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 *iqr* as a percentage of the median. This suggests that, *ceteris paribus*, even the best examples of these targeting methods produces relatively small targeting gains. By contrast, while other methods report higher median values, they are also characterized by proportionately higher variations in targeting effectiveness. So while these methods offer potentially large gains, there is no guarantee that they will improve targeting performance.

One way of exploring the source of variation in targeting outcomes is by using a Theil inequality index. A desirable feature of the Theil index is that it is sub-group decomposable; by grouping our data by some characteristic, we can allocate variation in targeting across these programs into two categories: that due to variations within each group and that due to variations across groups. When programs are grouped by region, we find that variation in average performance across continents explains only about 28 percent of total variation. Grouping according to program type, we find that variation in average performance between programs explains 36 percent of the total variation. Grouping by targeting method (according to whether they use geographic, means/proxy means, both, or other targeting methods) explains only 20 percent of the total variation.

One way of interpreting these large variations is in terms of implementation effectiveness. No matter how well one chooses among methods or programs, effectiveness of implementation is a key factor in determining targeting performance. This point is further illustrated 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 the average targeting performance from 1.35 to 1.49, a return of 14 percentage points. In section 4, we return to this issue but first it is necessary to note several caveats to be borne in mind when interpreting these findings.

3.4 Caveats and limitations

There are a number of caveats and limitations that should be made explicit with regard to interpreting our performance measure and, thus, the analysis based on it.
First, our performance measure is a mish-mash of various measures as discussed above, although for the vast majority of the interventions (80 percent) we use the percentage of

benefits accruing to either the bottom 40 percent or 20 percent of the national income distribution. This raises concerns regarding comparability. For example, one may believe that it is more difficult to target the poorest 20 percent compared to the poorest 40% so that programs for which we use the former may appear ineffective solely because of the performance indicator used. We have addressed this issue in a number of ways. We calculated a second performance measure that gives, through its lexicographic ordering, priority to the proportion of resources flowing to the bottom decile, then bottom quintile, then bottom two quintiles. Doing so does not change in any meaningful way the results reported in Tables 3 and 4. We also ran all regressions (reported below) using both measures of targeting performance and again found no meaningful change to our results. This is not completely surprising given that our performance measure and the alternative have correlation coefficients (in terms of levels and ranks respectively) between 0.94 and 0.97. As a further check, in the multivariate regression analysis we always include variables that control for the performance measure used.

Second, by focusing on the percentage of benefits accruing to the bottom parts of the income distribution we are obviously ignoring 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 those just above this income cut-off. This is partly why we give priority to the 40 percent measure of performance when constructing our performance index. It is also arguably the case that such a focus coincides more closely to the objectives of most targeted programs. In any case, the fact that our results are extremely insensitive to the ordering is at least suggestive that where between 20 percent and 40 percent one draws the cut-off point is somewhat inconsequential.

Third, recall that the data we have collated are only a sample of the hundreds of anti-poverty interventions that exist. Further, we could only calculate our performance indicator for two-thirds of this sample. These observations when taken together point to the possibility of "sample selection bias", that is to say that there may be certain characteristics of these programs – for example, the fact that they were evaluated and documented – which are themselves associated with our measures of targeting performance. A good example of this possibility relates to community targeting. Our sample is only a fraction of the studies listed in Conning and Kevane (2001); it could well be that only successful interventions using community targeting have been well documented.

Fourth, some of the mis-targeting 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. Methods that rely on static indicators of living standards (such as proxy means tests) are likely to perform less well than those that rely on self-selection when there is considerable movement of households in and out of poverty.

We remind the reader that we have been able to focus on only one narrow piece of the targeting and program choice decisions. Our performance index focuses solely on the benefit side of the equation and ignores cost, and the latter may be an extremely important factor in choosing targeting methods or programs to transfer income to the poor. For example, it is often argued that well-designed public works programs can be very effective at concentrating benefits in the hands of the poor. However, the high non-transfer costs associated with such programs (including non-wage costs and. forgone income) substantially reduces the cost-effectiveness of such transfer programs in this regard. Our ignoring of the cost side largely reflects data restrictions. In conducting the literature review we collated the available evidence on administrative costs, hoping to comment on how these varied by method. Unfortunately, such data were scant. We have some sort of cost data for 32 programs, but both cost and our performance indicator 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. These numbers were based on budget or expenditure records for program administration and thus include only official costs. No attempt is made to determine how much of program benefits are siphoned off due to corruption or theft. In contrast much of the cost data on South Asian programs is constructed from knowing a total budget and having data from a survey sample on the value of benefit received by households. Through appropriate grossing up, a figure for the total cost per dollar of benefit received is calculated. In most cases it appears that corruption and theft contribute more to total program expenses than legitimate administrative expenses, though little is said about these latter. 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, while always important, is often only one of the objectives of interventions. Therefore, to the extent that there are trade-off between these other objective and that of effective targeting – earlier we pointed to the possibility of social and political costs - this needs to be taken into account when arriving at an overall evaluation of any program. However, it may be the case that these other objectives impinge as much, if not more so, on the program design, the targeting process, and the way in which the program is "sold" and delivered. Presumably most policy analysts would at least accept that monitoring the targeting performance of programs dedicated mainly to poverty alleviation is always desirable, especially in the context of developing countries where poverty is high, budgets are tight, and other policy instruments (e.g. an comprehensive income tax system) are less developed, less sophisticated and less progressive.

4. Regression analysis

Although factors other than choice of method or program may be relatively large, this does not mean that these choices are unimportant. To get an idea of the importance of these choices, Table 5 presents the results of a series of regressions which identify how performance varies systematically across these choices. In doing so, we note that targeting methods are themselves choices, they are not "exogenous" or "pre-determined". Consequently, it is incorrect to treat these results as causal relations. Rather, they are measures of partial correlation or association.

Our first specification explores how these choices are associated with (log) incidence. We include dummy variables for nine targeting methods described above: three forms of individual assessment (means testing, proxy means testing, community selection of individual beneficiaries), four forms of categorical targeting (geographic, the elderly, the young, and others), and two types of selection (work requirement, community bidding for projects). The omitted category is self-selection based on consumption. We chose this as the base category for two reasons. It is often argued that this form of targeting should be seen as a transition tool while the capacity for more precise mechanisms – such as means testing – is developed.[7] Conversely, others have expressed skepticism over the ability of alternative targeting methods to reach the poor when compared to self-selection based on the consumption of food.[8] Hence, an attractive feature of this specification is that one should interpret the coefficients on these methods relative to self-selection based on consumption. We also include, but do not report, controls indicating whether the performance measure is based on the proportion of benefits going to the bottom quintile, the poorest decile, the "poor" defined with reference to a poverty line or the proportion of poor found in population. Doing so takes into account confounding effects arising from the use of different measures of incidence in the studies on which this analysis is based. Standard errors are computed using the methods proposed by Huber (1967) and White (1980).

Specification (1) 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 two quintiles relative to self-selection based on consumption. Proxy means testing, community assessment, targeting the young are also associated with improved incidence, though these are measured with larger standard errors. Targeting the elderly, other types of categorical targeting and selection based on community bidding are not associated with better incidence relative to our base category, self-targeting based on consumption.

Countries with better capacity for program implementation may do better at directing benefits towards poorer members of the population either by choosing finer targeting methods or implementing their choices more effectively. As such, the associations in specification (1) may be misleading; they may merely reflect correlation between unobserved implementation capacity and observed targeting methods. We explore this possibility in specifications (2) (3) and (4).

In specification (2) we include log GDP per capita (in PPP dollars) as of 1995 as an additional regressor. The hypothesis is that as a country becomes wealthier, it acquires the institutional capacity needed to design a well-targeted intervention. The positive and significant coefficient on income is consistent with such an argument. While the inclusion of income does not appear to reduce the coefficients on means testing or geographical targeting, coefficients on proxy means testing, community assessment and targeting the young effectively fall to zero and remain imprecisely measured. Selection

[7] See, for example, Pinstrup-Andersen (1988) and Alderman and Lindert (1998).
[8] Such implicit concern is found, for example, in Cornia and Stewart (1995).

based on a work requirement is still associated with improved incidence relative to selection based on consumption, but the coefficient is considerably smaller.

Specification (3) explores further the issue of implementation capacity by including measures of voice and governance found in Kaufmann, Kraay and Zoido-Lobaton (1999). Compiling subjective perceptions regarding the quality of governance in different countries using sources such as polls of experts, commercial risk rating agencies and cross-country surveys, they define voice, perhaps more accurately described as 'voice and accountability', as a composite measure based on aspects of political processes, civil liberties and political rights and thus captures the extent to which citizens participate in the selection of their governments as well as the extent to which citizens and media can hold governments accountable for their actions. Government effectiveness combines perceptions of the quality of public service provision, the competence of civil servants and the credibility of governments' commitment to policies. We use countries' percentile ranks (their ranking relative to each other) as these provide an easier way of interpreting the estimated coefficients. At 6, Viet Nam has the lowest percentile rank for "voice" while Costa Rica has the highest percentile rank, 88. Uzbekistan obtains the poorest governance rank at 6; Chile the highest rank at 86. In addition, we include country-specific Gini coefficients on the grounds that because targeting requires variation across individuals, it is plausible that identifying potential beneficiaries is easier when differences across individuals are greater.

Controlling for governance, voice and inequality 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 inequality and where governments are held accountable for their actions. Conditional on country income, better governance does not improve targeting but these latter two variables are highly correlated. When we drop log income in specification (4), we find that targeting is better in countries with better governance and voice. To give a sense of the magnitude of these effects, raising governance rank from 30 (the rank reported for Nicaragua) to 73 (the rank reported for Costa Rica) would be associated with an improvement in targeting performance of 0.29 or about a 30 per cent improvement relative to neutral targeting. Raising the voice rank from 37 (Pakistan's voice rank) to 67 (India's voice rank) would be associated with a similar improvement in targeting performance.[9]

We performed three additional specific checks to investigate the robustness of this result. Specifications (5), (6) and (7) use the same set of controls as specification (4) but restrict the data by the manner in which the performance indicator is measured. Specification (5) only includes studies that report the share of benefits accruing to the bottom two quintiles. Specification (6) includes studies that report the share of benefits accruing to the bottom two quintiles, or if that datum is not available, the share of benefits going to the poorest quintile. Specification (7) includes studies that report the share of benefits

[9] 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.

accruing either to the bottom two quintiles, the poorest quintile or the poorest decile. Where more than one measure is available, we use the measure relating to the larger number of individuals (so if shares going to the bottom 20 percent and 40 percent are both available, we use shares going to the bottom 40 percent). As we expand the sample in this way, the coefficient on geographic targeting increases, the coefficient on targeting based on a work requirement falls and the coefficient on means testing stays about the same but all three are positive and significant. Across specifications (2) through (7), means testing is associated with improvements in targeting performance, relative to self-selection based on consumption, of 21 to 27 percent. Geographic testing and self-selection based on a work requirement improve targeting by 20 to 36 per cent. By contrast, when we only look at studies reporting shares going to the bottom two quintiles, the coefficient on targeting the elderly, while negative, is not statistically significant whereas targeting children is associated with an improvement in targeting performance. However, the magnitudes and precision of these two coefficients does appear sensitive to the construction of the dependent variable. As we widen the sample, the negative impact of targeting the elderly increases in magnitude (and becomes more precisely measured) while the positive coefficient on targeting to young children disappears.

Specification (8) uses the same sample and regressors as specification (4), but the dependent variable is expressed in levels instead of logs. Our basic results remain unchanged: means testing and geographic targeting raise targeting performance relative to the omitted category, self-selection based on consumption. The coefficient on targeting based on a work requirement rises markedly, but is less precisely measured and there is no meaningful change in any of our other results.

Specification (9) takes a slightly different approach, estimating median regressions, which express differences in performance in terms of differences in medians.[10] 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. Relative to specification (4) – which uses an identical set of regressors, sample and dependent variable – the median regression reports larger coefficients on means testing and geographic targeting, a larger (though less precisely measured) coefficient on targeting via a work requirement, and a negative coefficient on targeting the elderly. The only change is that targeting to the young is now associated with improved targeting performance.

Our 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. We have not explored the association between combinations of targeting methods and targeting performance despite the fact that use of multiple methods is common. Table 6 remedies this omission. In addition to controls for income, voice, governance, inequality and how the performance measure is constructed, we include in specification (1) the number of targeting methods used. The results show

[10] More precisely, we estimated a quantile regression centred 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.

that use of more methods is associated with improved targeting, each additional method improves performance by 18 per cent. In specification (2), we represent the number of targeting methods by a series of dummy variables. This produces a similar finding. Unfortunately, our sample size is too small to explore the association between specific groupings of methods and targeting performance but these results are suggestive tha t such an approach improves targeting.

5. Conclusion

This paper addresses the contested issue of the efficacy of targeting interventions in developing countries using a newly constructed database of 111 targeted anti-poverty interventions found in 47 countries. We use these data to address three questions: 1) What targeting outcomes are observed? 2) Are there systematic differences in targeting performance by method and other factors? 3) What are the implications for such systematic differences for the design and implementation of targeted interventions?

We find that the median value of our measure of targeting performance is 1.25, so that the median program transfers 25 percent more to the target group than would be the case with a universal (or random) allocation. In this sense, “targeting works”. However, a staggering 21 of the 77 programs for which we can build our performance measure– more than a quarter -- are regressive, with a performance index less than one. In these cases, a random selection of beneficiaries would actually 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. However, even when these are dropped from our sample, we still find that 16 per cent of targeted anti-poverty interventions are regressive.

Countries with better capacity for program implementation, as measured either by GDP per capita or indicators of “governance” do better at directing benefits towards poorer members of the population. Countries where governments are more likely to be held accountable for their behaviour – where “voice” is stronger – also appear to implement interventions with improved targeting performance. Targeting is also better in countries where inequality is more pronounced and presumably differences in economic wellbeing are easier to identify.

Mindful of the caveats enumerated in section 3.4, interventions that use means testing, geographic targeting, and self-sele ction based on a work requirement are all associated with an increased share of benefits going to the bottom two quintiles relative to self-selection based on consumption. Demographic targeting to the elderly, community bidding, and self-selection based on consumption show limited potential for good targeting. Proxy means testing, community-based selection of individuals and demographic targeting to children show good results on average, but with considerable variation. That said, we again emphasize that there is considerable variation in targeting performance when we examine experiences with specific program types and specific targeting methods. Indeed a Theil decomposition of the variation in outcome shows that differences between targeting methods account for only 20 percent of overall variation, the remainder is due to differences found within categories. Thus it is not surprisinging that while community assessment generally performs no better relative to 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 as we have shown here, targeting the elderly generally performs relatively poorly when compared to other methods for reaching the poor.

Thus, while 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 appropriate targeting decisions. This suggests that further 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 paper (Coady, Grosh and Hoddinott (2002a), we provide a more detailed discussion of the merits and limitations of individual targeting methods in an attempt to move in this direction.

Table 1: The distribution of interventions by region and country income levels

	Transfers			**Subsidies**		**Public works for**	
	Cash	**Near cash**	**Food**	**Food**	**Non-food**	**Job creation**	**Program output (e.g. social funds)**
By Regions							
Latin America and Caribbean, *28*	12	3	3	0	1	4	5
Eastern Europe and FSU, *24*	22	1	0	0	0	0	1
Middle East and North Africa, *13*	0	0	0	13	0	0	0
Sub-Saharan Africa, *12*	3	0	1	4	1	2	1
South Asia, *21*	1	13	3	0	0	4	0
East Asia, *13*	4	1	4	0	2	1	1
By Income Level							
Poorest, *58*	14	15	6	9	2	7	5
Less poor, *53*	28	3	5	8	2	4	3
Total, *111*	**42**	**18**	**11**	**17**	**4**	**11**	**8**

Notes. 1. Numbers in italics are number of interventions by region and income level. 2. Poorest countries have per capita GDP in PPP dollars below 1200, less poor countries have per capita GDP above 1200 and below 10840.

Table 2: The distribution of targeting methods by region, country income levels and program type

	Individual assessment			**Categorical**				**Self selection**		
	Means tests	**Proxy means tests**	**Community assessment**	**Geography**	**Age - elderly**	**Age - children**	**Other**	**Work**	**Con-sumption**	**Community bidding**
By region										
Latin America and Caribbean, *65*	6	4	3	19	4	13	6	4	0	6
Eastern Europe and FSU, *41*	12	1	3	1	6	10	7	0	0	1
Middle East and North Africa, *17*	3	0	0	1	0	0	0	0	12	1
Sub-Saharan Africa, *21*	3	0	2	3	4	1	1	2	4	1
South Asia, *45*	2	1	3	16	2	1	6	4	10	0
East Asia, *37*	3	1	3	8	4	7	8	1	0	1
By income level										
Poorest, *129*	10	3	10	34	8	12	21	7	18	6
Less poor, *97*	19	4	4	14	12	20	8	4	8	4
By program type										
Cash transfer, *87*	19	4	5	8	15	21	15	0	0	0
Near-cash transfer, *36*	4	3	0	12	1	2	4	0	10	0
Food transfer, *30*	0	0	5	9	3	8	4	0	0	1
Food subsidy, *21*	3	0	0	1	0	0	0	0	16	1
Non-food subsidy, *8*	3	0	0	1	1	1	2	0	0	0
Public works, job creation, *26*	0	0	2	9	0	0	4	11	0	0
Public works, program output (eg social fund), *18*	0	0	2	8	0	0	0	0	0	8
***Total*, 226**	**29**	**7**	**14**	**48**	**20**	**32**	**29**	**11**	**26**	**10**

Notes. 1. Many programs use more than one targeting method. Thus the total number of targets methods tallied is greater than the number of programs. 2. Poorest countries have per capita GDP in PPP dollars below 1200, less poor countries have per capita GDP above 1200 and below 10840.

Table 3: Targeting performance by anti-poverty intervention

			Income level		Individual assessment			Categorical				Self-selection		
Country	Program Type	Performance	< 1200	>1200 & <10840	Means test	Proxy means test	Community assessment	Geographic	Age – elderly	Age - children	Other	Work	Consumption	Community bidding
Argentina	PW	4.00		✍				✍				✍		
Estonia	CT	3.47		✍	✍									
Hungary	CT	2.72		✍	✍									
Albania	CT	2.65	✍				✍				✍			
Poland	CT	2.10		✍	✍						✍			
Chile	CT	2.08		✍		✍				✍	✍			
Nicaragua	CT	2.02	✍					✍		✍	✍			
Honduras	CT	1.99	✍					✍		✍	✍			
Chile	FT	1.98		✍		✍	✍	✍		✍				
Slovenia	CT	1.95		✍	✍					✍				
Bolivia	PW	1.93	✍					✍				✍		
Chile	CT	1.83		✍		✍			✍					
Peru	FT	1.80		✍				✍		✍				✍
Chile	PW	1.78		✍								✍		
Indonesia	NFS	1.68	✍					✍			✍			
Bulgaria	CT	1.65		✍	✍									
India	NCT	1.63	✍		✍			✍					✍	
Mexico	NCT	1.60		✍	✍			✍		✍				
India	NCT	1.58	✍					✍					✍	
Hungary	CT	1.57		✍						✍				
Mexico	CT	1.56		✍		✍		✍		✍				
Costa Rica	FT	1.55		✍				✍		✍				
Colombia	NFS	1.50		✍	✍									
Indonesia	PW	1.48	✍					✍			✍	✍		
Costa Rica	CT	1.48		✍	✍				✍					
Jamaica	NCT	1.45		✍						✍	✍			
Indonesia	CT	1.44	✍				✍	✍		✍	✍			
India	NCT	1.36	✍					✍					✍	
Zambia	NFS	1.35	✍		✍									
Uzbekistan	CT	1.35	✍		✍		✍			✍	✍			

Table 3 continued

			Income level		Individual assessment			Categorical				Self-selection		
Country	Program Type	Performance	< 1200	>1200 & <10840	Means test	Proxy means test	Community assessment	Geographic	Age – elderly	Age - children	Other	Work	Consumption	Community bidding
Latvia	CT	1.33		✓						✓				
India	NCT	1.33	✓					✓					✓	
Indonesia	NCT	1.32	✓			✓		✓			✓			
Bolivia	SF	1.30	✓					✓						✓
Jamaica	NCT	1.30		✓	✓				✓		✓			
Romania	CT	1.29		✓						✓				
Honduras	SF	1.25	✓					✓						✓
Chile	CT	1.25		✓		✓								
India	NCT	1.25	✓					✓					✓	
Sri Lanka	NCT	1.25	✓		✓									
S. Africa	FS	1.23		✓									✓	
Vietnam	FT	1.22	✓					✓	✓	✓	✓			
India	NCT	1.20	✓					✓					✓	
Bangladesh	FT	1.20	✓				✓	✓		✓	✓			
Morocco	FS	1.18		✓									✓	
India	NCT	1.13	✓					✓					✓	
Armenia	CT	1.13	✓							✓				
Peru	SF	1.10		✓				✓						✓
Bulgaria	CT	1.10		✓					✓					
Nicaragua	SF	1.10	✓					✓						✓
India	NCT	1.09	✓		✓			✓					✓	
Zambia	SF	1.08	✓					✓						✓
Mozambique	CT	1.05	✓		✓				✓	✓	✓			
India	NCT	1.04	✓					✓					✓	
Tunisia	FS	1.03		✓									✓	
Uzbekistan	CT	1.01	✓		✓		✓				✓			
Egypt	FS	1.00	✓										✓	
India	NCT	1.00	✓					✓					✓	
Latvia	CT	1.00		✓	✓				✓	✓				
Egypt	FS	0.98	✓										✓	

Table 3 continued

			Income level		Individual assessment			Categorical				Self-selection		
Country	Program Type	Performance	< 1200	>1200 & <10840	Means test	Proxy means test	Community assessment	Geographic	Age – elderly	Age - children	Other	Work	Consumption	Community bidding
Bulgaria	CT	0.95		✍						✍				
Egypt	FS	0.95	✍		✍								✍	
Egypt	FS	0.95	✍		✍								✍	
Armenia	SF	0.93	✍					✍						✍
Tunisia	FS	0.93		✍									✍	
Poland	CT	0.90		✍	✍					✍				
Romania	CT	0.90		✍					✍		✍			
Morocco	FS	0.85		✍									✍	
S. Africa	FS	0.79		✍									✍	
Latvia	CT	0.70		✍					✍					
Algeria	FS	0.70		✍									✍	
S. Africa	FS	0.68		✍									✍	
Morocco	FS	0.60		✍									✍	
Armenia	NCT	0.58	✍			✍								
Yemen	FS	0.45	✍										✍	
Vietnam	CT	0.40	✍						✍		✍			
S. Africa	FS	0.28		✍									✍	

Notes:

1. CT: cash transfer; FS: food subsidy; FT: food transfer; NCT: near-cash transfer; NFS: non-food subsidy; PW: public works; SF: social fund.

Table 4: Targeting performance by targeting method

Targeting method	Sample size	Median targeting performance	Interquartile range	Interquartile range as percentage of median
All methods	77	1.25	0.56	44.8
Any form of individual assessment	30	1.40	0.73	52.1
Means testing	20	1.35	0.61	45.2
Proxy means testing	6	1.44	0.58	40.3
Community assessment	6	1.40	0.78	55.7
Any categorical method	53	1.32	0.50	37.9
Geographic	31	1.33	0.51	38.3
Age – elderly	10	1.08	0.40	37.0
Age – young	22	1.45	0.60	41.4
Other categorical	18	1.40	0.79	56.4
Any selection method	36	1.10	0.38	34.5
Work	4	1.85	1.34	72.4
Consumption	25	1.00	0.35	35.0
Community bidding	7	1.10	0.22	20.0

Table 5: Multivariate analysis of targeting performance

	Basic results				Robustness checks based on sample restrictions			Use level as dependent variable	Use median regression
	(1)	(2)	(3)	(4)	Studies reporting benefits to poorest 40% (5)	Studies reporting benefits to poorest 40 or 20%, (6)	Studies reporting benefits to poorest 40, 20 or 10%, (7)	(8)	(9)
Means testing	0.215 (2.30)**	0.218 (2.52)**	0.240 (3.18)**	0.278 (3.48)**	0.236 (2.79)**	0.245 (3.09)**	0.216 (2.63)**	0.242 (2.53)**	0.405 (2.95)**
Proxy means testing	0.203 (1.01)	-0.019 (0.12)	-0.110 (0.74)	0.009 (0.06)	-0.031 (0.21)	-0.074 (0.47)	-0.117 (0.75)	-0.194 (1.08)	-0.200 (0.68)
Community assessment	0.138 (0.69)	0.046 (0.24)	-0.098 (1.00)	0.052 (0.54)	-0.062 (0.69)	-0.062 (0.58)	-0.065 (0.56)	-0.125 (0.90)	-0.253 (0.95)
Geographic	0.252 (2.83)**	0.352 (3.84)**	0.327 (3.65)**	0.215 (2.93)**	0.225 (2.12)**	0.309 (3.15)**	0.353 (3.51)**	0.391 (3.23)**	0.511 (3.21)**
Age – elderly	-0.140 (0.96)	-0.184 (1.47)	-0.221 (1.90)*	-0.293 (2.41)**	-0.150 (0.86)	-0.238 (1.92)*	-0.238 (1.90)*	-0.236 (1.81)*	-0.280 (1.09)
Age – young	0.177 (1.68)*	0.033 (0.34)	0.036 (0.37)	0.087 (0.89)	0.175 (1.83)*	0.109 (1.45)	0.011 (0.10)	0.0001 (0.00)	0.266 (1.70)*
Other categorical	-0.005 (0.05)	0.187 (1.72)*	0.229 (2.27)**	0.194 (1.78)*	0.331 (3.46)**	0.222 (2.52)**	0.224 (2.22)**	0.222 (1.33)	0.174 (0.81)
Work	0.571 (2.88)**	0.285 (1.81)*	0.230 (1.67)*	0.359 (2.52)**	0.384 (2.92)**	0.288 (2.08)**	0.200 (1.34)	0.496 (1.44)	0.493 (0.92)
Community bidding	-0.049 (0.49)	-0.049 (0.44)	-0.092 (0.82)	-0.003 (0.03)	0.038 (0.39)	-0.051 (0.44)	-0.119 (0.91)	-0.215 (1.31)	-0.231 (1.07)
Log GDP per capita		0.246 (4.43)**	0.211 (3.07)**		0.068 (0.89)	0.175 (2.46)**	0.227 (2.84)**	0.282 (3.20)**	0.194 (1.81)*
Voice			0.005 (2.67)**	0.007 (3.61)**	0.003 (1.91)*	0.005 (3.01)**	0.005 (2.59)**	0.005 (2.09)**	0.004 (1.30)
Governance			-0.0004 (0.20)	0.005 (2.22)**	0.004 (1.68)*	0.0006 (0.27)	-0.0008 (0.32)	-0.0009 (0.35)	0.0002 (0.06)
Inequality			0.009	0.008	0.006	0.008	0.009	0.018	0.013

			(2.19)**	(1.94)*	(1.43)	(1.72)*	(1.92)*	(2.51)**	(1.65)*
F statistic	2.77**	4.80**	6.55**	5.41**	9.65**	7.86**	6.86**	5.95**	
R2	0.427	0.553	0.648	0.596	0.713	0.721	0.674	0.651	
Sample size	77	77	76	76	48	60	64	76	76

Notes:

1. Specifications (1) – (4), (8) and (9) contain controls, not reported, indicating whether performance measure is based on proportion of benefits going to the (a) bottom quintile, (b) poorest decile, (c) to the “poor” or (d) proportion of poor found in population. Specification (7) contains only (a) and (b), specification (6) contains only (a) and specification (6) contains no controls.
2. Specifications (1) – (8) estimate standard errors using the methods proposed by Huber (1967) and White (1980). Specification (9) calculates standard errors using the bootstrap with 50 repetitions.
3. Specifications (1) –(7) and (9) express the dependent variable in logs; specification (8) uses levels.

Table 6: Association between targeting performance and number of methods used

	(1)	(2)
Number of methods used	0.137 (3.38)**	
Used two methods		0.110 (1.12)
Used three methods		0.293 (2.89)**
Used four methods Log GDP per capita	 0.189 (2.75)**	0.372 (3.01)** 0.185 (2.65)**
Voice	0.005 (2.93)**	0.005 (2.66)**
Governance	-0.002 (0.65)	-0.002 (0.62)
Inequality	0.013 (2.98)**	0.013 (2.82)**
F statistic	6.03**	5.02**
R2	0.506	0.508
Sample size	76	76

Notes:

1. Specifications (1) and (2) contain controls, not reported, indicating whether performance measure is based on proportion of benefits going to the (a) bottom quintile, (b) poorest decile, (c) to the "poor" or (d) proportion of poor found in population.
2. Specifications (1) and (2) estimate standard errors using the methods proposed by Huber (1967) and White (1980).

References

Ahmad, E. and N. Stern. 1991. *The theory and practice of tax reform in developing countries.* Cambridge: Cambridge University Press.

Alderman, H. 2002. "Do Local Officials Know Something We Don't? Decentralization of Targeted Transfers in Albania". *Journal of Public Economics* 83(3): 375-404.

Alderman, H. and K. Lindert. 1998. "The potential and limitations of self-targeted food subsidies." *World Bank Research Observer* 13(2): 213-229.

Atkinson, A. 1995. "On Targeting Social Security: Theory and Western Experience with Family Benefits". In D. van de Walle and K. Nead, eds., *Public spending and the poor.* Baltimore: Johns Hopkins University Press.

Besley, T. and R. Kanbur. 1993. "Principles of targeting." In M. Lipton and J. van der Gaag, eds., *Including the Poor*. Washington DC: World Bank.

Bigman, D. and H. Fofack. 2000. "Geographic targeting for poverty alleviation: An introduction to the special issue." *World Bank Economic Review* 14(1): 129-146.

Braithwaite, J. C. Grootaert, and B. Milanovic. 2000. *Poverty and Social Assistance in Transition Countries*. New York: St. Martin's Press.

Case, A. and A. Deaton. 1998. "Large cash transfers to the elderly in South Africa." *Economic Journal*, 108(4): 1330-1361.

Coady, D., and E. Skoufias. 2001."On the Targeting and Redistributive Efficiencies of Alternative Transfer Programs." FCND Discussion Paper 100. International Food Policy Research Institute, Washington, DC.

Coady, D., M. Grosh and J. Hoddinott. 2002a. "The targeting of transfers in developing countries: Review of experience and lessons." Mimeo. International Food Policy Research Institute, Washington, DC.

Coady, D., M. Grosh and J. Hoddinott. 2002b. "Targeted anti-poverty interventions: A selected annotated bibliography." Mimeo. International Food Policy Research Institute, Washington, DC.

Coady, D., R. Perez and H. Vera-Llamas. 2001. "Cost-Effectiveness Analysis of the Education Component of PROGRESA." In D. Coady, ed., *The Application of Social Cost-Benefit Analysis to the Evaluation of PROGRESA*, Final report submitted to PROGRESA, International Food Policy Research Institute, Washington D. C.

Conning, J. and M. Kevane. 2001. “Community Based Targeting Mechanisms for Social Safety Nets.” Social Protection Discussion Paper No. SP 0102. World Bank, Washington, D.C.

Cornia, G. and F. Stewart. 1995. “Two errors of targeting” In D. van de Walle and K. Nead, eds., *Public spending and the poor*. Baltimore: Johns Hopkins University Press.

Datt, G. and M. Ravallion. 1994. “Transfer Benefits From Public-Works Employment: Evidence For Rural India.” *Economic Journal*, 104: 1346-69.

Duflo, E., 2000. “Child Health and Household Resources in South Africa: Evidence from the Old Age Pension Program.” *American Economic Review*, Papers and Proceedings, 90(2): 393-398.

Grosh, M. 1994. *Administering Targeted Social Programs in Latin America: From Platitudes to Practice*.” Washington DC: World Bank.

Huber, P. 1967. “The behavior of maximum likelihood estimates under non-standard conditions.” in *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*. University of California Press, Berkeley CA.

Jalan, J. and M. Ravallion. 1999. “Income gains to the poor from workfare: Estimates for Argentina’s Trabajar Program.” Policy Research Working Paper 2149. World Bank, Washington D.C.

Kaufmann, D., A. Kraay and P. Zoido-Lobaton. 1999. “Aggregating Governance Indicators.” Policy Research Working Paper 2195. World Bank, Washington, D.C.

Newbery, D. and N. Stern, eds.1987. *The theory of taxation for developing countries* Oxford University Press, New York.

Pinstrup-Andersen, P., ed., 1988. *Food Subsidies in Developing Countries: Costs, Benefits and Policy Options*. Baltimore: Johns Hopkins University Press.

Ravallion, M. 1993. “Poverty Alleviation through Regional Targeting: A Case Study for Indonesia.” K. Hoff, A. Braverman and J. E. Stiglitz, eds., *The Economics of Rural Organization: Theory Practice and Policy*. New York: Oxford University Press.

Ravallion, M. and K. Chao. 1989. “Targeted Policies for Poverty Alleviation Under Imperfect Information: Algorithms and Applications.” *Journal of Policy Modeling*, 2(2): 213-224.

Ravallion, M. and G. Datt. 1995. “Is Targeting through a Work Requirement Efficient?” In D. van de Walle and K. Nead, eds., *Public spending and the poor*. Baltimore: Johns Hopkins University Press.

Rawlings, L., L. Sherburne-Benz and J. Van Domelen. 2001. "Letting Communities Take The Lead: A Cross-Country Evaluation of Social Fund Performance" (Draft), The World Bank, Washington, D.C.

van de Walle, D. 1998. "Targeting revisited." *World Bank Research Observer* 13(2): 231-248.

White, H. 1980."A heteroscedasticity-consistent covariance matrix and a direct test for heteroscedasticity." *Econometrica*, 48: 817-838.

World Bank (2000) *World Development Report: Attacking poverty,* Oxford University Press, New York.

World Bank (1997) *World Development Report: The role of the state,* Oxford University Press, New York.

World Bank (1990) *World Development Report: Poverty,* Oxford University Press, New York.

WB17929
A:\jnl_jh7final.doc
November 8, 2002 2:46 PM