How Relevant Is Targeting to the Success of an Antipoverty Program?

Martin Ravallion

Policy-oriented discussions often assume that "better targeting" implies larger impacts on poverty or more cost-effective interventions for fighting poverty. The literature on the economics of targeting warns against that assumption, but evidence has been scarce and the lessons from the literature have often been ignored by practitioners. This paper shows that standard measures of targeting performance are uninformative or even deceptive about the impacts on poverty, and cost-effectiveness in reducing poverty, of a large cash transfer program in China. The results suggest that in program design and evaluation, it would be better to focus directly on the program's outcomes for poor people than to rely on prevailing measures of targeting. JEL codes: I32, I38, O15

Like many developing countries, China is trying to develop an effective system of social protection based on cash transfers targeted to the poor. The largest such program that China has attempted—and almost certainly the largest such program in the developing world—is the "Minimum Livelihood Guarantee Scheme," popularly known as *Di Bao*. This has been the government's main response to the new challenges of social protection in urban areas. While economic reforms and structural changes in the Chinese economy have meant high rates of economic growth, it is believed that certain sub-groups have been adversely affected or have been unable to participate in the new economic opportunities due to long-term illness, disability, or their lack of skills. The collapse of the old safety net provided by guaranteed employment has clearly left some households vulnerable. Some of the "left behind" households started poor and some became poor, even though aggregate poverty rates have tended to fall over time. Urban areas have figured prominently in these concerns about the "new poor."

The *Di Bao* program aims to address this new social problem. The scheme became a national policy in 1999 and expanded rapidly; by 2003 participation had leveled off at 22 million people, representing 6 percent of urban residents. As is typically the case in developing countries (and many developed countries), the transfers made under the program are targeted to the income poor. There are many ways of implementing such targeting in practice, including a direct means test, a proxy-means test, and self-targeting mechanisms such as those based on work requirements.[1] *Di Bao* relies heavily on a direct means test, whereby potential recipients report their income and if this is below a stipulated "poverty line" for that locality then a transfer payment is made to bring that family up to the line. There are local-level checks on eligibility. Municipal authorities have considerable power over the program, including setting the *Di Bao* lines, funding (the center provides partial co-financing), and implementation.

Naturally, China's policy-makers and the community at large want to know whether such a large and ambitious public welfare program is in fact achieving its objective. Does it reach the intended beneficiaries and how much impact does it have on poverty? Various measures of the "targeting performance" have been widely used to help address that question. These measures are typically interpreted by both analysts and policy-makers as indicators of a program's performance in "directing benefits toward poorer members of the population" (Coady, Grosh, and Hoddinott 2004a, p. 81). Comparisons of such measures across different programs have informed public choices on which programs should be scaled up and which should be dropped.[2]

As in any situation in which measurement is used to inform policy, "the indicators need to be related to the overall policy problem, with an explicit formulation of the objective and constraints" (Atkinson 1995, p. 31). It is widely agreed that the objective of this class of public programs is to reduce poverty subject to the relevant resources and constraints, including those related to the information available and the behavior of relevant agents. Better targeting is not seen as being desirable in its own right, but rather as an instrument for reducing poverty.

Do the measures of targeting used in policy discussions provide useful indicators for this policy problem? The most widely used measures quantify some aspect of how well a given program concentrates its benefits on the poor, which is essentially what "targeting" has come to mean. An example is the share of transfers going to the poor. Cornia and Stewart (1995) have been influential in arguing that measurement practices and policy discussions have put too high a weight on avoiding one type of error—the Type 1 error of having ineligible non-poor participants—in order to address the Type 2 error of incomplete coverage of the poor.[3]

Cornia and Stewart did not present data linking these aspects of targeting performance to poverty outcomes (though they did point to this as an important direction for further research). However, the literature warns us against assuming

that better targeting, as assessed by standard measures, will necessarily enhance a program's total impact on poverty.[4] A number of factors cloud the relationship between targeting performance and total impact on poverty, including aspects of program design, implementation, and the context in which a program operates. Incentive issues have been a theme of one strand of the literature, pointing to the possibility that fine targeting will impose high marginal tax rates on recipients, possibly creating poverty traps.[5] The literature has also warned that fine targeting can undermine political support for an antipoverty program; concentrating gains on the poor may induce a lower overall transfer to the poor, with benefits spread too thin or covering too few people.[6]

It is also unclear how useful these measures are as indicators of cost-effectiveness and as an input to revising policy decisions. Here the focus is not on the total impact, but rather on the impact per unit of the resources devoted to a given program. (The total impact then depends on the allocation of resources across programs, weighted by their cost-effectiveness ratios.) Intuitively, the impact on poverty will depend on both the share of transfers going to the poor and the total transfer. Plainly a large uniform transfer (received by everyone, whether poor or not) can have more impact on poverty than a small well-targeted transfer. But will the latter type of program, with low leakage to the non-poor, necessarily be more cost-effective? The answer is far from obvious on *a priori* grounds. The factors noted above which cloud the relationship between targeting performance and a program's total impact on poverty will not, in general, vanish when total impact is normalized by total spending.

For example, finer targeting typically entails administrative costs, which are debits against the total budget in determining the government's total transfer payment. Then the share of transfers going to the poor does not even identify the transfer to the poor per unit of public spending. Less obviously, but no less importantly, targeting can generate hidden costs to participants, notably when there are restrictions such as work requirements, behavioral conditions, or sources of social stigma. Given the costs of targeting, it is not difficult to imagine cases in which the better targeted program (with the higher share of transfers going to the poor) is less cost-effective in reducing poverty: the literature already contains examples.[7] In short, avoiding leakage to the non-poor can reduce the amount actually going to the poor, with theoretically ambiguous implications for poverty and cost-effectiveness in fighting poverty.

Whether better targeting, as measured in practice, implies a greater impact on poverty or a more cost-effective intervention is ultimately an empirical question. Yet, beyond a few suggestive examples, we really know rather little about how well these popular targeting measures perform in practice.

This paper tries to help fill this gap in knowledge using a case study of China's *Di Bao* program. A number of factors—the decentralized nature of the program,

its scale, and the availability of a large data set representative at local level—combine to make this an unusual opportunity to put targeting measures to the test. The program's targeting performance and impacts on poverty are estimated under standard assumptions across each of the 35 major municipalities of China. The most popular targeting measures found in policy-oriented discussions are thus tested as indicators of program performance in reducing poverty.

Measuring Targeting Performance and Poverty Impacts

In principle, one might choose to measure "targeting performance" by a program's impact on poverty relative to an explicit comparison, such as an untargeted allocation of the same budget (as in Ravallion and Chao, 1989). Then the interpretation for poverty is unambiguous. That is not, however, the approach that has dominated the literature and practice. This discussion will focus on the main measures of targeting performance found in practice, on which much of our current knowledge about what does and does not work is based. More precise definitions of the measures can be found in table 1.[8]

Targeting Measures

The present discussion will focus on four main measures. The first three measures are based on the concentration curve $[C(p)]$, which gives the cumulative share of transfers to the poorest percentage of the population as ranked by, for example, household income per person; figure 1 gives a hypothetical concentration curve. If the transfers under a given program are uniform—in that each person receives the same amount—then the concentration curve is simply the 45 degree diagonal line.[9] Intuitively, an actual curve farther from the diagonal indicates better targeted transfers made by a given program.

   The first of the four measures is the share of transfers going to the poorest percentage, defined as H. This is simply one point on the concentration curve, namely $S = C(H)$. In the empirical work discussed later, it will be natural to identify H as the target group for the program: that is, the set of people deemed to be has incomes below the municipal *Di Bao* poverty line. More precisely, we can set $H = H_0$, which is the pre-intervention headcount index of poverty—the proportion of the population living in households with pre-transfer income per person less than the poverty line. (The post-transfer headcount index is denoted as H_1.) For much of the present discussion we can just take H to be some reference group of poor or relatively poor people, without presuming that it is the precise target population for the program in question.

Table 1. Measures of Targeting and Poverty

| Measure	Definition	Formula
Targeting measures		
Concentration curve ($C(p)$)	Share of total transfers going to the poorest percentage of the population (p) ranked by household income per person.	$C(p) = (1/\bar{t}) \int_0^p t(x)\, dx$
Share going to the poor (S)	Share of transfers going to those who are initially deemed poor (or other reference group based on income).	$S = C(H_0)$
Normalized share (NS)	Share going to the poor divided by proportion who are poor.	$NS = C(H_0)/H_0$
Concentration index (CI)	Area between the concentration curve (above) and the diagonal (along which everyone receives the same amount).	$CI = 2 \int_0^1 C(p)\, dp - 1$
Coverage rate (CR)	Program participation rate for the poor.	$CR = N(D=1, Y < Z)/N(Y < Z)$
Targeting differential (TD)	Difference between the coverage rate and the participation rate for the non-poor.	$TD = N(D=1, Y < Z)/N(Y < Z) - N(D-1, Y \geq Z)/N(Y \geq Z)$
Proportion of Type 1 errors ($T1$)	Proportion of ineligible non-poor who are assigned the program.	$T1 = N(D=1, Y \geq Z)/N(Y \geq Z) = (1-S)P/(1-H_0)$
Proportion of Type 2 errors ($T2$)	Proportion of the poor who fail to receive the program.	$T2 \equiv N(D=0, Y < Z)/N(Y < Z) = 1 - CR$
Poverty measures		
Headcount index (H)	The proportion of the population living in households with income per person less than the poverty line.	$H = F(Z)$
Poverty gap index (PG)	Mean distance below the poverty line as a proportion of the line where the mean is taken over the whole population, counting the non-poor as having zero gap.	$PG = \int_0^H (1 - y(p)/Z)\, dp$
Income-gap ratio (I)	Mean distance below the poverty line as a proportion of the line, amongst the poor alone.	$I = 1 - PG/H$

Table 1. *Continued*

Measure	Definition	Formula
Squared poverty gap index (*SPG*)	As for PG except that the proportionate poverty gaps are weighted by themselves.	$SPG = \int_0^H (1 - y(p)/Z)^2 \, dp$
Poverty impact	Pre-intervention poverty measure minus post-intervention measure	$H_0 - H_1$; $PG_0 - PG_1$; $SPG_0 - SPG_1$

Note: The table only refers to the measures used in this paper. The text contains further discussion of the pros and cons of each measure. Notation: $t(x)$ is the transfer to quantile x ranked by income per person and $\bar{t}$ is the mean transfer; H_0 is the pre-program headcount index of poverty (and similarly for *PG* and *SPG*); H_1 is the post-intervention headcount index (and similarly for *PG* and *SPG*); $N(D = 1, Y < Z)$ is the number of people who are both poor ($Y < Z$) and receiving the transfers; $N(Y < Z)$ is the number of poor; $N(D = 1, Y \geq Z)$ is the number of people who are both non-poor and receiving the transfers; $N(Y \geq Z)$ is the number of non-poor; $P = N(D = 1)/N$ is the participation rate; F is the cumulative distribution function (CDF) for incomes; Z is the poverty line; and $y(p)$ is the quantile function (inverse of CDF) giving income of the pth qantile.

Figure 1. Targeting Measures Based on the Concentration Curve
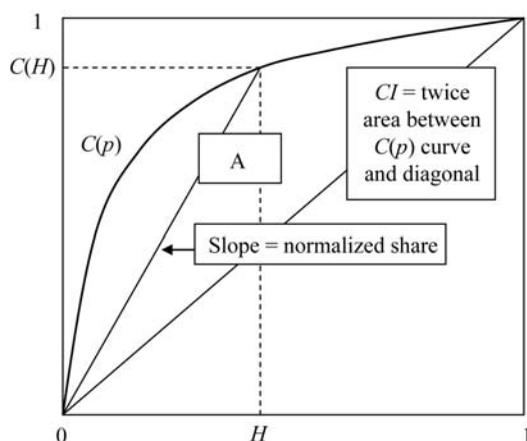
Note: see table 1 for notation.

The popularity of the concentration curve (S) is evident in the fact that the meta-studies by Grosh (1994, 1995) and Coady, Grosh, and Hoddinott (2004a, 2004b) found that this was the most readily available measure in their many primary sources.[10] The measure's popularity may well stem from its ease of interpretation. Against this advantage, the measure has some obvious drawbacks. For one thing, it tells us nothing about how transfers are distributed amongst the poor; two programs can have the same share of transfers going to the poor, but in one case the gains are heavily concentrated amongst the poorest whereas in the other they only reach those just below the poverty line. Another concern is that this measure does not directly reflect the overall size of the transfer program, which will clearly matter to impacts on poverty, as discussed in the introduction.[11]

The second measure is the normalized share (NS) obtained by dividing S by H (figure 1 shows NS). Coady, Grosh, and Hoddinott (2004a, 2004b) preferred to use this as their measure of targeting performance, arguing that this was more comparable than the ordinary share (S) because it measures performance relative to a "common reference outcome...that would result from neutral (as opposed to progressive or regressive) targeting" (p. 69).[12] By "neutral targeting" they mean a uniform transfer. If the transfer is uniform then clearly NS equals one. However, finding a value of NS "close" to unity does not imply that the allocation is "close" to being uniform. There are many ways one could get a value for NS near unity, with rather different interpretations. Similarly to S, the NS measure is insensitive to how transfers are distributed amongst the poor. The poor can receive their population share (H) of the transfers, but different people amongst the poor receive very different amounts. For example, the money could all go to

either the poorest person or the least poor person; either way NS would equal one. NS also approaches unity as H approaches 100 percent, no matter how the money is distributed. When the reference outcome is this ambiguous, the usefulness of the measure becomes theoretically questionable.

The third measure is the concentration index (CI), which is widely used in studies of fiscal incidence. This can be thought of as a "generalized S" in that, instead of focusing on one point on the concentration curve, CI measures the area between the curve and the diagonal (along which the transfer is uniform); in figure 1, CI is just twice the area marked A.[13] The index is bounded above by 1 (at which point the poorest person receives all payments) and below by -1 (the richest person receives all). This measure has the attraction that it reflects distribution amongst the poor and even over the whole range of incomes. A disadvantage is that it is not as easy to interpret as S or NS. The same CI can entail very different allocations of transfers. And, as with the previous measures, it tells us nothing directly about the scale of the transfers.

Although these measures are all based on the concentration curve, they can give quite different results. Of course, S and NS will always be in the same ratio to each other when the same value of H is used for all programs. However, these two measures can rank programs differently when H varies, both as in the case study presented later in this paper and presumably in many applications. To illustrate, consider a transfer scheme operating in two cities and giving all participants the same amount. In city A all the transfers go to the poorest 20 percent and the overall poverty rate is 50 percent: in city B the transfers go to the poorest 40 percent and the poverty rate is 10 percent. A far higher share of the transfers goes to the poor in A ($S = 100$ percent) versus in B ($S = 25$ percent). City A also has the higher concentration index ($CI = 0.8$ in A versus 0.6 in B). In contrast, it is in city B where the scheme is deemed to be better targeted according to the normalized share ($NS = 2.5$ for B versus 2 for A). More generally, the concentration curve for program A would lie entirely above that for program B and yet NS is higher for B, given its lower H.

The fourth measure is the targeting differential (TD), which is the difference between the program's participation rate for the poor—which one can call the coverage rate—and that for the non-poor (table 1).[14] Alternatively, one can normalize the targeting differential by the mean transfer over all recipients (TD^*). When the transfer is uniform, $TD = TD^*$. It turns out that the choice between TD and TD^* makes very little difference in this case study. Since TD is easier to interpret the present discussion shall focus on this measure.

To interpret the targeting differential, note that when only the poor get help from the program and all of them are covered, $TD = 1$, which is the measure's upper bound; when only the non-poor get the program and all of them are covered, $TD = -1$, its lower bound. (In the example of two cities above, $TD =$

0.67 for city B and 0.4 for A.) This measure is easy to interpret, and it automatically reflects both leakage to the non-poor and coverage of the poor.

How are these measures related to the incidence of Types 1 and 2 errors as discussed in the introduction? A Type 1 error can be defined as incorrectly classifying a person as poor, whereas a Type 2 error is incorrectly classifying a person as not poor. A Type 1 error entails a leakage of transfers to the non-poor, whereas a Type 2 error implies lower coverage of the poor. In measuring the proportions of Types 1 and 2 errors, one can normalize by the populations of the non-poor and poor (respectively);[15] see table 1 for more precise definitions. Standard targeting measures depend on the incidence of both types of errors. It should not be presumed that measures based on the concentration curve will be largely unaffected by Type 2 errors; indeed, all these measures can be thought of as functions of the proportions of these two types of targeting errors.[16] That is also true of the targeting differential, for which the relationship is particularly clear: *TD* is simply 1 minus the total proportions of Types 1 and 2 errors.[17] So this particular measure automatically gives equal weight to both types of errors. However, for the measures based on the concentration curve that weights are attached to these two errors of targeting is an empirical question.

Poverty Impacts

In testing the relevance of these targeting measures to this case study's poverty impacts, three poverty measures will be used: the headcount index, the poverty gap index (*PG*), and the squared poverty gap index (*SPG*; introduced by Foster, Greer, and Thorbecke 1984). The measures are defined in table 1. The pros and cons of each are well documented; for a review see Ravallion (1994). Briefly, *H* is the easiest to interpret and is the most popular measure, but is unaffected by income gains or losses to the poor unless they cross the poverty line. *PG* reflects mean income of the poor, but not inequality amongst the poor, which is the main advantage of *SPG*.

Impacts are measured by pre-transfer minus post-transfer poverty measures. Impacts on these measures are estimated on the same data and under the same assumptions about how the scheme works (including behavioral responses), as used in measuring targeting performance. In particular, it will be assumed that income in the absence of the program is observed income minus payments received under the *Di Bao* program. This assumes that there is no displacement of other income sources through behavioral responses, such as reduced work effort or lower private transfer receipts. This is the most common assumption in the literature on measuring targeting performance; indeed, it appears that virtually all of the primary studies used by Coady, Grosh, and Hoddinott (2004a, 2004b) made this assumption. The assumption is questionable, although the paper will offer some tests that suggest the data are at least consistent with the assumption.

In assessing cost-effectiveness, the poverty impacts are normalized by the cost of the program, though a more flexible econometric method of controlling for total spending will also be used. Given the costs of targeting, it is not difficult to imagine cases in which the better targeted program by any of the above measures is less cost-effective against poverty. Consider again the example of cities A and B above, in which the program in city A is better targeted according to both S and CI but not NS or TD. Suppose that the total cost to the government is the same, but that the finer targeting of city A's program (for which it will be recalled that all of the transfers go to the poorest 20 percent versus 40 percent in city B) entails extra costs to both the government and participants such that only 25 percent of participants in city A escape poverty, whereas in B all poor participants are able to do so. The headcount index falls by 5 points in A, but 10 points in B. B's program has a higher impact on poverty and is more cost-effective.

There is a special case in which one of these measures, namely the share S, is a perfect indicator of cost-effectiveness for PG. That special case is when the program has no impact on H and there are no fiscal costs besides the transfers. Then it can be readily shown that the impact on PG per unit public spending is simply S. Of course, this special case is unlikely to be of much practical interest, given that people in a neighborhood of the poverty line will presumably be transfer recipients and there will undoubtedly be other costs.

Under the same assumptions, it can be readily shown that the normalized share (NS) is a perfect indicator of cost-effectiveness in reducing the income-gap ratio, as given by the mean income gap of the poor as a proportion of the poverty line (table 1). This is (implicitly) the poverty measure relevant to comparisons of program performance based on the normalized share. However, as a poverty measure the income-gap ratio is known to have a number of undesirable properties; for example, if a poor person living above the mean for the poor escapes poverty then this measure perversely suggests higher poverty. PG does not have this property.

It should be noted that these measures of poverty impacts and cost-effectiveness can all be calculated from the same data required for the various measures of targeting performance described above. Of course, if one knows the impacts on poverty—which we agree to be the objective—then one does not need the targeting measures. However, since these targeting measures are widely used in assessing antipoverty programs and in comparative work, it is of interest to test their value as indicators for that policy problem.

Data on China's *Di Bao* Program

This case study uses China's Urban Household Short Survey (UHSS) for 2003–2004, as discussed in Chen, Ravallion, and Wang (2006). The UHSS was done by

the Urban Household Survey Division of the National Bureau of Statistics (NBS). The paper uses the UHSS sample for the 35 largest cities with a total sample of 76,000, varying from 450 in Shenzhen to 12,000 in Beijing. For these 35 cities, the definitions of geographic areas in the UHSS coincide with those for the *Di Bao* lines and the entire data set has been cleaned by NBS staff and made available for this research.

While the UHSS is a relatively short survey, its results allow us to measure a fairly wide range of household characteristics. The survey also included a question on household income and questions were added on *Di Bao* participation and income received from the program.

The UHSS data confirm one's expectation that China's cities vary in ways that could well be relevant to the outcomes of the *Di Bao* program. For example, across the 35 largest urban areas studied in this paper, the highest mean household income per person (the city of Shenzhen) is over four times than that of the lowest (Chongqing). The proportion living below the *Di Bao* poverty line varies from 2 percent (in Fuzhou) to 19 percent (Haikou). As noted in the last section, in measuring targeting and poverty impacts it is assumed that income in the absence of the program is observed income minus the reported payments received under the *Di Bao* program. While this is a common assumption, it is clearly questionable. Testing the assumption is difficult without panel data (and even then there can be severe identification problems). With only a single cross-sectional survey it is hard to be confident in the results, given the likelihood of omitted variables correlated with both program placement and the behaviors of interest.

However, there are some observations that are at least consistent with this assumption. The design of the *Di Bao* program intends that the benefits received will decrease as income rises, implying that participants face a positive marginal tax rate. Indeed, if the program works the way it is supposed to then it exactly fills the gap between current non-*Di Bao* income and the *Di Bao* line. Then participants will have no incentive to work (under the usual assumptions that leisure is a normal good and work yields no direct utility). Earned income net of transfers from *Di Bao* will fall to zero. The program will have created a poverty trap, whereby participants do not face an incentive to raise their own incomes, because of the loss of benefits under the program.

The extent to which this is a real problem in practice is unclear. Benefits are unlikely to be withdrawn quickly. There are reports that local authorities allow *Di Bao* benefits to continue for some period after the participant finds a job (O'Keefe 2004). Observations from field work also indicate that a notion of "imputed income" was used in a number of provinces. This was a notional level of income that reflected the potential income given the household labor force; this was apparently done with the aim of minimizing work disincentives.[18]

Chen, Ravallion, and Wang (2006) studied how the *Di Bao* payment per capita varied with the *Di Bao* gap, given by the difference between the relevant poverty line and income net of the program's transfer (both per capita). If the program exactly filled these gaps when positive then *Di Bao* payments would rise with a slope of unity, but would be zero for those with income above the *Di Bao* line. Chen, Ravallion, and Wang found a marked tendency for mean transfer payments (conditional on the *Di Bao* gap) to rise with the *Di Bao* gap, though the conditional expected value as measured by a non-parametric regression has a slope appreciably less than unity. The regression line starts to be noticeably positive at per capita incomes that are about 2,000 Yuan above the *Di Bao* line and then peaks at a mean of around 300 Yuan per capita, at a *Di Bao* gap of around 4,000 Yuan. (The conditional mean is, of course, positive throughout, but very close to zero below 2,000 Yuan.) Thus the average benefit withdrawal rate—the amount by which mean transfer payments change with an extra Yuan of income—is around −0.05; on average, a 100 Yuan increase in income entails a drop of only 5 Yuan in transfer payments.

An alternative method of estimating the average benefit withdrawal rate is to regress the per capita *Di Bao* payment received on income per person minus the *Di Bao* receipts, with a complete set of dummy variables for municipalities (to capture the differences in the generosity of the program). The implied benefit withdrawal rate is very low, at −0.0012 (*t*-ratio = −17.51, *n* = 76,808).[19]

There is almost certainly attenuation bias in these estimates due to income measurement errors. There is the usual source of measurement error in reporting incomes using only one question, amplified by the fact that income net of transfer payments will probably underestimate income in the absence of the program if there are behavioral responses. To address this concern, one can use an instrumental variables estimator, in which a set of household-level characteristics (including demographics, education attainments, occupation, and housing conditions) are used as instrumental variables for income in estimating the benefit withdrawal rate; Chen, Ravallion, and Wang (2006) provide details on the variables used in the first-stage regressions. Note that this only works for the unconditional regression coefficient of *Di Bao* payments on pre-transfer income, so the instrumental variables are automatically excluded from the main regression of interest; the conditional benefit withdrawal rate is unidentified. The instrumental variables estimate of the unconditional benefit withdrawal rate is −0.0021 (*t*-ratio = −28.33), again very low. These calculations were repeated separately for each municipality, using the instrumental variables estimator for the full sample in each municipality. The estimates were significantly negative for all municipalities and ranged from −0.0102 to −0.0001.

While each of these tests requires an assumption that can be questioned, they all suggest that the benefit withdrawal rate for *Di Bao* is very small. It would thus

Table 2. Leakage and Coverage of the *Di Bao* Program

Percent of population	Net income below Di Bao poverty line		Total
	Yes	*No*	
Receiving *Di Bao*	2.22	1.69	3.91
Not receiving *Di Bao*	5.48	90.60	96.09
Total	7.71	92.29	100.00

Note: $n = 76,443$ for the 35 municipalities.

appear unlikely that the program would provide any serious disincentive for earning income, thus supporting our assumption that income in the absence of the program is simply observed income minus the transfer payments received. At the same time, such a low benefit withdrawal rate raises concerns about how well the program reaches the poorest and how well it adapts to changes in household needs. For a stylized version of this type of transfer program, Kanbur, Keen, and Tuomala (1995) find that an optimal benefit withdrawal rate around one-half is consistent with evidence on the relevant income elasticity of labor supply.

Targeting Performance and Poverty Impacts of the *Di Bao* Program

On calculating all these measures on the same data set and under the same assumptions, one can test the assumption commonly made in policy discussions that better targeting allows a greater impact on poverty and/or a more cost-effective antipoverty program. One can also revisit some of the findings from past research on the factors relevant to targeting success. The discussion begins with the aggregate results and then turns to the city-level analysis.

Aggregate Results

Across the 35 cities, 7.7 percent of the total population had a net income (observed income minus *Di Bao* receipts) below the *Di Bao* line (table 2). The program's total participation is equivalent to about half of the eligible population. About 40 percent of program recipients are ineligible according to these data ($0.43 = 1.69/3.91$). The proportion of these Type 1 errors amongst the non-poor is clearly very low at 0.018 ($1.69/92.29$). But there is a high proportion of Type 2 errors, with almost three-quarters of those who are eligible not being covered by the program ($0.71 = 5.48/7.71$, i.e., a coverage rate of 29 percent).[20] Nonetheless, targeting performance

Table 3. Impacts on Aggregate Poverty Measures for Urban China

	Poverty measures (percent)	
	Before Di Bao (income net of receipts)	After Di Bao (income including receipts)
(a) Population (participants + non-participants)		
Headcount index (percent)	7.71	7.26
Poverty gap index (percent)	2.28	2.06
Squared poverty gap index ($\times 100$)	1.02	0.88
(b) Participants only		
Headcount index (percent)	56.85	45.49
Poverty gap index (percent)	19.92	14.23
Squared poverty gap index ($\times 100$)	10.21	6.44

appears to be excellent by international standards, as judged by the three measures based on the concentration curve; one finds that $S = 64$ percent, $NS = 8.3$ and $CI = 0.78$.[21] Coady, Grosh, and Hoddinott (2004a, 2004b) provide estimates of NS for 85 programs. Argentina's *Trabajar* program has an $NS = 4.0$, making it the best performer by this measure amongst all programs surveyed by Coady, Grosh, and Hoddinott.[22] The median NS is 1.25. By this measure, *Di Bao* is a clear outlier in targeting performance internationally.

Turning to the fourth measure of targeting performance we find that while 29 percent of the poor receive a transfer under the program, this is only true of about 2 percent of the non-poor. Thus the targeting differential (TD) is 0.27. The mean transfer payment across all those with net income below the *Di Bao* line is 87.61 Yuan per person per year, while the corresponding mean for those with income above the line is 4.15. The overall mean *Di Bao* payment across all recipients is 270.33, so when the targeting differential is normalized by the mean transfer one obtains $TD^* = 0.31$.

These calculations indicate that, while the program is good at concentrating benefits on the poor, it still falls well short of the perfect targeting ($TD = 1$) that would cover all of the poor and only the poor as implied by the program's design. Another way to see this is to calculate the total receipts for those with net income below the *Di Bao* line. One then finds that only 12.1 percent of the aggregate *Di Bao* gap is filled by the program. The program is a long way off reaching its own aim of bringing everyone up to the *Di Bao* poverty line.

The weak coverage of the program—in terms of both coverage of those living below the *Di Bao* line and coverage of the *Di Bao* gap—is naturally limiting its impact on poverty, despite excellent targeting in the sense of avoiding leakage to the non-poor. Table 3 gives various poverty measures before and after the transfers received from the program.[23]

The program is having a sizeable impact on poverty amongst the participants (table 3). The proportion of the participant population falling below the *Di Bao* line is 45 percent with *Di Bao* transfers, but it would have been 57 percent without them. However, the impact on poverty in the population as a whole is much less. The proportion falling below the *Di Bao* lines falls from 7.7 percent to 7.3 percent after the transfers made under the program. Proportionate impacts are slightly higher for *PG* than for the headcount index (and slightly higher again for the squared poverty gap); this indicates that the program has increased the mean income of those below the *Di Bao* line and reduced inequality amongst them.

Targeting and Poverty across Cities of China

Given the scheme's decentralized financing and implementation, and the differences observed across China's cities, heterogeneity in outcomes across municipalities is to be expected. There will, of course, be differences in local resources and administrative capabilities, but there will also be less obvious differences in the local political economy. Here the aim is to describe the differences in program performance across municipalities, and to use these differences to assess how well prevailing targeting measures perform in predicting impacts on poverty.

There is considerable variation in targeting performance across municipalities.[24] *S* varies from 31 percent to 98 percent; *NS* varies from 2.8 to 18.8; *CI* varies from 0.64 to 0.93; and *TD* varies from 0.06 to 0.53. (All cities except one, Kunming, have *NS* higher than the best performing program surveyed by Coady, Grosh, and Hoddinott 2004a, 2004b.)

Recall that *TD* automatically gives equal weights to the Types 1 and 2 error proportions. The weights for the three measures based on the concentration curve depend on the analytic properties of these measures and how the design features and setting influence the overall program participation rates (as discussed above). A simple way of summarizing this potentially complex relationship is to regress each measure of targeting performance on the proportions of Types 1 and 2 errors. Table 4 gives the regressions. The proportion of Type 2 errors has only a small and statistically insignificant effect on the three targeting measures based on the concentration curve (*S*, *NS*, or *CI*). In contrast, the proportion of Type 1 errors has a strong and significant effect for the share and the normalized share; the coefficient is also significant at the 3 percent level for the share if the proportion of Type 2 errors is dropped.[25]

The normalizations used in defining the proportion of Types 1 and 2 errors clearly affect the results. If one normalizes by total populations of the municipality the results change noticeably.[26] As can be seen in table 4, the three measures based on the concentration curve all attach negative weights to Type 1 errors per capita, but now we find that *NS* also puts a negative weight on Type 2 errors (as

does *CI*, although it is not significant). And we find that *S* puts a positive weight on Type 2 errors, while *TD* puts a positive weight on Type 1 errors. Only for the concentration index do we find that the weights attached to the two errors of targeting are robust to the normalization. These findings lead to questioning of the generalizations found in the literature about the relative importance of Types 1 and 2 errors to standard measures of targeting.

However, the most important point for the present analysis is that none of the measures based on the concentration index have strong correlations with the coverage rate of the poor. The simple correlation coefficients with the coverage rate are -0.28, -0.30, and -0.40 for *S*, *NS* and *CI* respectively. In contrast, the targeting differential has a correlation of 0.98 with the coverage rate.

There are some clear covariates of the heterogeneity in targeting performance, echoing past findings in the literature. There is a high correlation between *Di Bao* spending and targeting performance as measured by *TD* ($r = 0.73$), though the correlation with targeting performance is appreciably weaker for the other measures. Indeed, the other targeting measures tend to be negatively correlated with program spending, though only significantly so for *CI* ($r = -0.52$). This pattern in the data is consistent with evidence for antipoverty programs in other settings indicating that *TD* tends to improve as programs expand and to deteriorate in fiscal contractions (Ravallion 2004). It appears that the early benefits at a low level of spending tend to be captured more by the non-poor, while the poor both benefit more when the program expands and are the first to bear the costs of contractions. The differences in program scale, as measured by participation rates, are also highly positively correlated with coverage rates for the poor ($r = 0.80$).

The impact of higher initial poverty on targeting performance has also been discussed in the literature.[27] The present results confirm the Coady, Grosh, and Hoddinott (2004a, 2004b) finding that the normalized share is higher in richer cities. The correlation coefficient between the normalized share and the pre-program headcount index is -0.81. However, this is entirely due to the normalization; if one uses the ordinary share (*S*) the correlation is positive and significant ($r = 0.55$).[28] (One wonders whether the Coady, Grosh, and Hoddinott results would be robust to using *S*, or *NS* with a uniform *H* across all programs.)

Impacts on poverty also vary across cities (see the working paper version, Ravallion 2007, for details). Subtracting the post-*Di Bao* poverty rate from the pre-*Di Bao* rate, the impact on the headcount index varies from 0.0 to 1.5 percentage points. Table 5 provides the correlation coefficients between the targeting measures and the program's impacts on both the headcount index and poverty gap index. The results for *SPG* were very similar to *PG* and are omitted for brevity. Correlations are given for both the levels impact (pre-*Di Bao* poverty measure

Table 4. Targeting Measures Regressed on Types 1 and 2 errors

	Share (S)		Norrmalized share (NS)		Concentration index (CI)		Targeting differential (TD)	
Constant	0.560	0.542	9.689	16.355	0.902	0.871	1	0.288
	(2.43; 0.02)	(10.97; 0.00)	(1.89; 0.07)	(13.54; 0.00)	(11.79; 0.00)	(27.71; 0.00)		(8.65; 0.00)
Proportion of Type 1 errors	−0.938	n.a.	−0.390	n.a.	−1.354	n.a.	−1	n.a.
	(−1.04; 0.30)		(−2.21; 0.03)		(−4.59; 0.00)			
Proportion of Type 2 errors	0.138	n.a.	0.015	n.a.	−0.052	n.a.	−1	n.a.
	(0.42; 0.67)		(0.21; 0.84)		(−0.49; 0.63)			
Type 1 errors per capita	n.a.	−1.044	n.a.	−0.672	n.a.	−1.382	n.a.	2.165
		(−1.81 0.08)		(−6.88; 0.00)		(−6.77; 0.00)		(3.34; 0.00)
Type 2 errors per capita	n.a.	2.728	n.a.	−0.903	n.a.	−0.153	n.a.	−1.000
		(4.76; 0.00)		(−5.16; 0.00)		(−0.35; 0.73)		(−2.72; 0.01)
R^2	0.097	0.412	0.141	0.639	0.369	0.365	1	0.449

n.a. = not applicable.

Note: *t*-ratios and probability values in parentheses, based on White standard errors; $n = 35$.

Table 5. Is Targeting Performance Correlated With Poverty Impacts?

Correlation coefficients	Impact on poverty measure		Proportionate impact on poverty (normalized by pre-transfer value)		Cost-effectiveness ratio (impact on poverty per unit spending)	
	Headcount index	Poverty gap index	Headcount index	Poverty gap index	Headcount index	Poverty gap index
Program spending per capita	0.80*	0.86*	0.66*	0.69*	0.04	−0.17
(a) Measures of targeting performance						
Share of spending going to the poor (S)	0.03	0.07	−0.26	−0.33	0.32	0.65*
Share going to the poor normalized by headcount index (NS)	−0.44*	−0.53*	0.04	0.12	−0.21	−0.42*
Concentration index (CI)	−0.40	−0.40	−0.19	−0.17	0.04	0.26
Targeting differential (TD)	0.61*	0.65*	0.63*	0.74*	0.13	−0.01
(b) Types 1 and 2 errors						
Proportion of Type 1 errors (T1)	0.63*	0.72*	0.44*	0.51*	−0.06	−0.11
Proportion of Type 2 errors (T2)	−0.66*	−0.71*	−0.63*	−0.75*	−0.09	0.03

*Indicates significance at 1 percent level.
Note: $n = 35$.

minus post-*Di Bao* measure) and the proportionate impact (normalized by the pre-*Di Bao* poverty measure).

Amongst the four measures of targeting performance, by far the strongest indicator of the impact on poverty is the targeting differential (*TD*). Strikingly, there is no sign of a positive correlation between the impacts on poverty and any of the three more popular measures, *S*, *NS*, and *CI*. For *S* the correlation coefficient with the impacts on the level of the headcount index is only 0.03, while for *CI* and *NS* the correlation coefficients with poverty impacts turn out to be negative. This switches for proportionate impacts, which are negatively correlated with *S* but virtually orthogonal to *NS*. Figure 2 plots the impacts on the level of the headcount index against the *NS* measures. We see clearly that municipalities with a higher normalized share going to the poor tended to have lower impacts on poverty. The targeting differential does not have this perverse property, as can be seen from figure 3, though even *TD* is far from being a good predictor of poverty impacts.

In table 5, the correlations are pair-wise. Instead, table 6 gives regressions of the poverty impacts (columns 1 and 3) on all four measures jointly. (The table also includes regressions that control for spending, which will be discussed below.) The targeting differential remains the strongest predictor of poverty

Figure 2. Relationship between the Normalized Share of Benefits Going to the Poor and the Impact on Poverty of the *Di Bao* Program
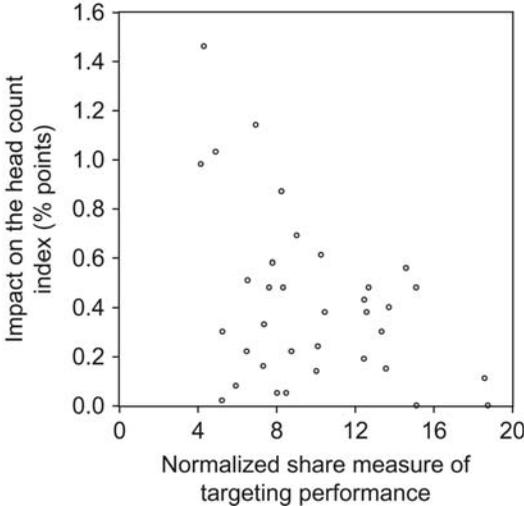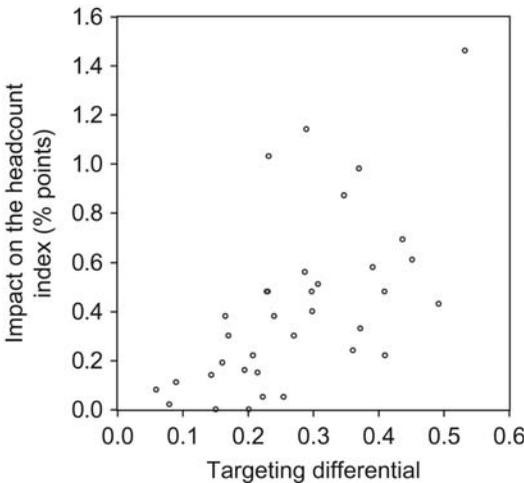

Figure 3. Relationship between the Targeting Differential and the Di Bao Program's Impact on Poverty

impacts. The share (S) now emerges as a positive predictor, though still not significant at the 1 percent level, and NS is no longer a significant negative indicator at given values of the other targeting measures. CI remains a negative predictor of poverty impacts.

Table 6. Which Targeting Measure Best Predicts Poverty Impacts and Cost-effectiveness?

	Impact on headcount index		Impact on poverty gap index		
	(1)	(2)	(3)	(4)	(5)
Constant	1.092 (1.59; 0.12)	0.401 (0.86; 0.40)	0.540 (1.78; 0.09)	0.017 (0.10; 0.92)	0.001 (0.00; 0.99)
Share of spending going to the poor (S)	0.722 (2.53; 0.02)	0.608 (2.56; 0.02)	0.413 (2.08; 0.05)	0.343 (2.49; 0.02)	0.338 (2.65; 0.01)
Normalized share going to the poor (NS)	−1.056 (−0.92; 0.37)	−0.306 (−0.46; 0.65)	−1.033 (−2.13; 0.04)	−0.706 (−1.77; 0.08)	−0.692 (−1.85; 0.07)
Concentration index (CI)	−1.800 (−2.02; 0.05)	−1.181 (−1.87; 0.07)	−0.881 (−1.80; 0.08)	−0.188 (−0.62; 0.54)	−0.183 (−0.67; 0.51)
Targeting differential (TD)	1.613 (4.61; 0.00)	−0.072 (−0.15; 0.88)	0.932 (5.92; 0.00)	0.177 (0.97; 0.34)	0.176 (1.04; 0.31)
Program cost	n.a.	1.099 (3.11; 0.00)	n.a.	0.125 (0.51; 0.62)	0.179 (4.65; 0.00)
Program cost squared	n.a.	−0.441 (−2.44; 0.02)	n.a.	0.044 (0.24; 0.81)	n.a.
Program cost cubed	n.a.	0.064 (2.51; 0.02)	n.a.	−0.007 (−0.25; 0.80)	n.a.
R^2	0.540	0.764	0.643	0.849	0.847

n.a. = Not applicable.

Note: t-ratios and probability values in parentheses, based on White standard errors; $n = 35$.

Table 5 also gives the correlation coefficients for the incidence of Types 1 and 2 errors. Strikingly, one finds that the proportion of Type 1 errors is positively correlated with poverty impacts; cities that did a better job excluding the non-poor tended to do less well in reducing poverty. It is only the incidence of Type 2 errors that is negatively correlated with impacts.[29] The same pattern is found in the partial correlations by regressing poverty impacts on both the proportions of Types 1 and 2 errors; the regression coefficients on the Type 1 error proportions were significantly positive at the 1 percent level for the impacts on both H and PG, whereas for Type 2 they were significantly negative.

It is now evident that the main reason why the three targeting measures based on the concentration curve perform so badly as indicators of poverty impact is that they are not even positively correlated with the program's coverage of the poor, which is highly correlated with poverty impacts (as can be seen from the correlation coefficients for Type 2 errors in table 5).

What about cost-effectiveness? Cities with higher *Di Bao* spending tend to have higher impacts on poverty; the correlation coefficients between *Di Bao* payments per capita and the impacts of H and PG are 0.80 and 0.86, respectively (table 5). The simplest way to test how well the targeting measures predict poverty impacts at given levels of spending is by normalizing poverty impacts by *Di Bao* spending to give the cost-effectiveness ratio.[30] Table 5 gives the correlation coefficients.

The share going to the poor now emerges as having a positive correlation with cost-effectiveness and being statistically significantly for PG. Figure 4 plots the data points in this case; a positive relationship is evident, with a regression

Figure 4. Cost-Effectiveness: Relationship between the Share Going to the Poor and Reduction of the Poverty Gap
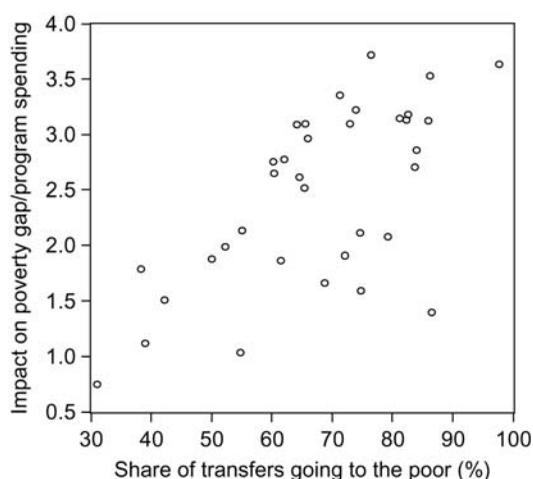
coefficient of 0.033 ($t = 5.61$, based on a White standard error). Even so $R^2 = 0.42$, so that the majority of the variance in cost-effectiveness in reducing the poverty gap index is left unexplained.

This is the exception though. None of the measures show significant correlations with cost-effectiveness in reducing the headcount index. The normalized share still has the perverse negative correlation found for the total poverty impacts.[31] The reason why the normalized share, NS, emerges as a perverse indicator (even though S is uncorrelated with impact and positively correlated with cost-effectiveness) is that there is a positive correlation between the pre-program headcount index and the program's impact on poverty; the correlation coefficients are 0.36 and 0.51 for H and PG, respectively. The incidence of the two types of errors shows little or no correlation with the cost-effectiveness ratios. (There is no sign of scale effects on those ratios, with correlation coefficients not significantly different to zero; see table 5.)

A more flexible way of seeing whether the targeting measures reveal poverty impacts from given spending is to use a regression of poverty impacts on the targeting measures, with controls for spending. To allow for nonlinearity in a reasonably flexible way, a cubic function of program spending per capita was used. Testing each targeting measure one by one, only S turns out to be a significant (positive) predictor of impacts on poverty from given program spending; the correlation was significant at the 4 percent level for H and 1 percent level for PG. The results using the four measures together are given in table 6 (columns 2 and 4; for PG the higher order terms can be dropped, giving column 5). Again, one finds that only S is a significant predictor of impacts on poverty from given program spending, though now this holds not only for PG at the 2 percent level, but also for H.

Conclusions

The three most popular measures of targeting performance found in practice are the share of transfers going to the poorest H percent, the share normalized by H, and the concentration index. Despite their popularity in analytic work and policy discussions on antipoverty programs, there has been little or no research into the performance of these measures in providing useful indicators for either the poverty impacts of such programs or their cost-effectiveness in reducing poverty. At the same time, the literature on the economics of targeting has repeatedly warned against assuming that a better targeted program—as judged by any of these measures—will have a greater impact on poverty.

This paper has provided some evidence. The results indicate that none of these measures reveal much about the success of China's *Di Bao* program in achieving

its objective of eliminating extreme urban poverty. The cities of China that are better at targeting this program, as assessed by these measures, are generally not the ones where the scheme came closest to attaining its objective. More encouragingly, one finds that a fourth measure proposed in recent literature—the targeting differential—does have a statistically significant positive correlation with the program's poverty impacts, although this is a relatively new measure which has not yet been widely used. But even this measure is far from being a perfect indicator of poverty impacts.

All these targeting measures are quite uninformative about the program's cost-effectiveness, shown as poverty impact from given program spending. The one exception is that the share going to the poor is a statistically significant predictor of cost-effectiveness in reducing the poverty gap index. But even then about 60 percent of the variance in the cost-effectiveness ratio is left unexplained. All other measures perform poorly or even perversely as indicators of cost-effectiveness.

These findings echo some of the warnings in the literature against relying on standard measures of targeting performance for informing policy choices on anti-poverty programs. The paper's findings also cast doubt on the generalizations found in the literature about what type of program works best (and should presumably be promoted) based on cross-program comparisons of targeting measures. The external validity of these programmatic comparisons is highly questionable when the targeting measures have such a poor fit with poverty impacts. It is also unlikely that past findings on the socio-economic factors influencing targeting performance at the country level are robust to seemingly arbitrary differences in the measures used.

One question is left unanswered: Why have the literature's warnings carried so little weight in practice? Possibly the more "theoretical" objections to these targeting measures have fallen on deaf ears for lack of clear evidence on how the measures perform in practice. The results of this case study will then help. One can also conjecture that the preference for targeting measures that put a high weight on avoiding leakage to the non-poor stems from fiscal pressures, given that reducing leakage helps cut public spending while expanding coverage does the opposite. While no doubt such thinking has had influence at times, it is surely misguided. For if the problem was to minimize public spending (unconditionally) then why would governments bother with such programs in the first place? Evidently there is a demand for these policies, as part of a comprehensive antipoverty strategy. A more credible characterization of the policy problem would then give positive weight to both avoiding leakage and expanding coverage of the poor.

From that perspective, measures of targeting performance that penalize both errors of targeting make more sense, again echoing recommendations found in the literature. However, that conclusion would still miss the point. Even the

targeting measure found here to be the best predictor of poverty impacts is a long way from being a perfect indicator. If there is a single message from this study it is that analysts and policy-makers would be better advised to focus on the estimable outcome measures most directly relevant to their policy problem. In the present context, impacts on poverty can be assessed with the same data and under the same assumptions as required by prevailing measures of targeting performance.

Funding

Notes

1. Overviews of the various methods found in practice can be found in Besley and Kanbur (1993), van de Walle (1998), Coady, Grosh, and Hoddinott (2004a, 2004b) and Ravallion (2006).

2. Early empirical studies by Mateus (1983) and Grosh (1992) were influential in arguing the case for finer targeting. The meta-studies of Grosh (1994) and Coady, Grosh, and Hoddinott (2004a, 2004b) have provided the most comprehensive comparative data on program performance drawn from targeting measures.

3. The distinction between these two errors goes back to Weisbrod (1970) who called them "vertical-" and "horizontal-target efficiency." Cornia and Stewart (1995) used the terms "E-mistakes" and "F-mistakes"; Smolensky, Reilly, and Evenhouse (1995) called them "errors of inclusion" and "errors of exclusion." The literature on social welfare policy in developed countries has also suggested that coverage of the poor is given little weight by standard measures of targeting. See, for example, the results of Duclos (1995) on the implications of incomplete enrollment in Britain's welfare benefits for measures of targeting. However, the relationship between these problems and overall impacts on poverty has received little attention.

4. For an overview of the arguments and evidence see van de Walle (1998).

5. Besley and Kanbur (1993) pointed to this problem and other issues raised by targeting. Also see Smolensky, Reilly, and Evenhouse (1995). Kanbur, Keen, and Tuomala (1995) study the incentive issues in fine targeting, including characterizing an optimal scheme for poverty reduction which takes account of labor supply responses.

6. For theoretical analyses see De Donder and Hindriks (1998) and Gelbach and Pritchett (2000).

7. Ravallion and Datt (1995) and Murgai and Ravallion (2005) provide examples for workfare programs in India.

8. Table 1 relates only to the measures used in this study. For a more comprehensive discussion of these and other measures, including their analytic properties, see the excellent volume by Lambert (2001).

9. This case is sometimes called an "un-targeted transfer" in the literature, although it is far from clear that the absence of any effort at targeting would yield a uniform transfer.

10. Coady, Grosh, and Hoddinott (2004a, 2004b) provide the shares going to the poorest 10 percent, 20 percent, and 40 percent for 85 of the antipoverty programs in their study (though with missing data in some cases).

11. The literature has pointed to the possibility that the share going to the poor can vary with the scale of a program, through the political economy of program capture (see Lanjouw and Ravallion 1999).

12. Coady, Grosh, and Hoddinott (2004a, 2004b) used $H = 40\%$ when it was available (which was the case for about half the programs in their study) and the next lowest available number (20 percent or 10 percent) when the value for $H = 40\%$ was not available. In the earlier comparative study of targeting performance by Grosh (1994), the value of H was set at 40 percent in all programs studied, in which case the first two measures will of course rank identically.

13. To assure that all measures go in the same direction, I multiply the usual definition of CI by -1.

14. This measure was proposed by Ravallion (2000). Also see Galasso and Ravallion (2005) on the properties of this measure, and the discussion in Stifel and Alderman (2005).

15. One might prefer to normalize by population size; similar formulae for this case are easily derived, but the essential point remains.

16. Consider, for example, the share (S). It is readily verified that $S = 1 - [T1(1 - H)/P]$ where P is the overall program participation rate ($T1$ is as defined in Table 1). Alternatively $S = 1 - T1^*$ where $T1^*$ is the proportion of participants who are Type 1 errors. But one can equally well write S as a function of Type 2 errors, namely $S = (1 - T2)H/P$ ($T2$ is as defined in Table 1) or $S = (H/P) - T2^*$. Nor is P likely to be independent of $T1$ and $T2$; for example, higher coverage of the poor (lower $T2$) may tend to come with larger programs. Thus S can be taken to depend on both $T1$ and $T2$.

17. More precisely, $TD = 1 - (T1 + T2)$ using the notation in Table 1.

18. This is based on a personal communication with Philip O'Keefe at the World Bank, drawing on his field-work discussions with local administrators.

19. This does not allow for censoring. Using a Tobit regression, the estimate is -0.004 (t-ratio $= -76.23$). Estimating the Tobits separately for each municipality, I obtained statistically significant benefit withdrawal rates in all cases, but all were very low and none were higher (in absolute value) than -0.001.

20. Ravallion (2008) proposes a method of testing robustness of such calculations to income measurement errors and implements the method for the same program. The results indicate that, if anything, the survey-based incomes probably understate the coverage of the eligible population.

21. To calculate CI from the micro data I used the regression-based method of Jenkins (1988).

22. *Trabajar* is a combination of a workfare program and social fund, whereby participants are offered low-wage work to do things of value to poor communities (see Jalan and Ravallion 2003). Coady, Grosh, and Hoddinott (2004a, 2004b) calculate the normalized share from *Trabajar* from Jalan and Ravallion (2003) who estimate that 80 percent of *Trabajar* participants come from the poorest 20 percent of the Argentine population ranked by income net of *Trabajar* receipts (which are roughly constant across recipients). The corresponding normalized share for the poorest decile is much higher, at about 6.0, though still less than for the *Di Bao* program.

23. To test robustness to the location of the poverty line, Chen, Ravallion, and Wang (2006) give the empirical cumulative distribution functions of income—with households ranked by income normalized by the relevant *Di Bao* line—with and without transfer receipts for both participants and the full 35-city sample. The qualitative results in Table 3 are robust to the choice of poverty line.

24. The working paper version gives detailed results by city (see Ravallion 2007).

25. The regression coefficient on Type 2 errors is then -1.361 (t-ratio $= -2.28$; prob $= 0.03$).

26. Thus the regressors become $H_0 T1$ and $(1 - H_0)T2$.

27. For a theoretical analysis using TD as the measure of targeting, see Ravallion (1999).

28. Using mean income instead, one obtains $r = 0.57$ for NS and -0.34 for S.

29. Expressing the incidence of targeting errors on a per capita basis gives virtually identical correlations for Type 1 errors. However, the per capita incidence of Type 2 errors is uncorrelated with poverty impacts. This stems from the strong positive correlation between (pre-*Di Bao*) H and poverty impacts (and the positive correlation between H and Type 2 errors per capita); when using a regression to control for H a significant negative correlation emerges between the poverty impacts and Type 2 errors.

30. Data were not available on administrative costs of the program at municipal levels, so the spending variable is solely based on transfer payments. The correlation coefficients will only be unaffected if the administrative cost share is the same across different cities.

31. The normalized share is significantly correlated with cost-effectiveness in reducing the income-gap ratio ($r = 0.53$), although it is necessary to recall that this is a flawed measure of poverty.

References

Atkinson, Anthony, B. 1995. "On Targeting and 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 Michael Lipton and Jacques van der Gaag, eds., *Including the Poor*. World Bank, Washington, D.C.

Chen, Shaohua, Martin Ravallion, and Youjuan Wang. 2006. "*Di Bao*: A Guaranteed Minimum Income in China's Cities?" Policy Research Working Paper, WPS 3805. World Bank, Washington, D.C.

Coady, David, Margaret Grosh, and John Hoddinott. 2004a. "Targeting Outcomes Redux." *World Bank Research Observer* 19(1):61–86.

Coady, David, Margaret Grosh, and John Hoddinott. 2004b. *Targeting Transfers in Developing Countries: Review of Lessons and Experience*. 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.

De Donder, Philippe, and Jean Hindriks. 1998. "The Political Economy of Targeting." *Public Choice* 95:177–200.

Duclos, Jean-Yves. 1995. "On the Equity Aspects of Imperfect Income Redistribution." *Review of Income and Wealth* 41(2):177–90.

Foster, James, J. Greer, and Erik Thorbecke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica* 52:761–65.

Galasso, Emanuela, and Martin Ravallion. 2005. "Decentralized Targeting of an Antipoverty Program." *Journal of Public Economics* 85:705–27.

Gelbach, Jonah, and Lant Pritchett. 2000. "Indicator Targeting in a Political Economy: Leakier Can Be Better." *Journal of Policy Reform* 4:113–45.

Grosh, Margaret. 1992. "The Jamaican Food Stamps Program: A Case Study in Targeting." *Food Policy* 17(1):23–40.

_____. 1994. *Administering Targeted Social Programs in Latin America: From Platitudes to Practice*. World Bank, Washington, D.C.

———. 1995. "Toward Quantifying the Trade-Off: Administrative Costs and Incidence in Targeted Programs in Latin America." In Dominique van de Walle and Kimberly Nead, eds., *Public Spending and the Poor*. Baltimore, MD: Johns Hopkins University Press.

Jalan, Jyotsna, and Martin Ravallion. 2003. "Estimating the Benefit Incidence of an Antipoverty Program by Propensity-Score Matching." *Journal of Business and Economic Statistics* 21(1):19−30.

Jenkins, Stephen. 1988. "Calculating Income Distribution Indices from Micro-Data." *National Tax Journal* 41(1):139−42.

Kanbur, Ravi, Michael Keen, and Matti Tuomala. 1995. "Labor Supply and Targeting in Poverty-Alleviation Programs." In Dominique van de Walle and Kimberly Nead, eds., *Public Spending and the Poor*. Baltimore, MD: Johns Hopkins University Press.

Lambert, Peter. 2001. *The Distribution and Redistribution of Income, Third Edition*. Manchester, UK: Manchester University Press.

Lanjouw, Peter, and Martin Ravallion. 1999. "Benefit Incidence and the Timing of Program Capture." *World Bank Economic Review* 13(2):257−74.

Mateus, Abel. 1983. *Targeting Food Subsidies for the Needy: The Use of Cost-Benefit Analysis and Institutional Design*. World Bank Staff Working Paper 617. Washington, DC: The World Bank.

Murgai, Rinku, and Martin Ravallion. 2005. "Is a Guaranteed Living Wage a Good Antipoverty Policy?" Policy Research Working Paper 3640. World Bank, Washington, D.C.

O'Keefe, Philip. 2004. "Social Assistance in China: An Evolving System." Mimeo. World Bank, Washington, D.C.

Ravallion, Martin. 1994. *Poverty Comparisons*. Chur, Switzerland: Harwood Academic Press.

———. 1999. "Are Poorer States Worse at Targeting their Poor?" *Economics Letters* 65:373−77.

———. 2000. "Monitoring Targeting Performance when Decentralized Allocations to the Poor are Unobserved." *World Bank Economic Review* 14(2):331−45.

———. 2004. "Who is Protected from Budget Cuts?" *Journal of Policy Reform* 7(2):109−22.

———. 2006. "Transfers and Safety Nets in Poor Countries: Revisiting the Trade-Offs and Policy Options." In Abhijit Banerjee, Roland Benabou, and Dilip Mookerjee, eds., *Understanding Poverty*. Oxford, UK: Oxford University Press.

———. 2007. "How Relevant is Targeting to the Success of an Antipoverty Program?" Policy Research Working Paper 4385. Washington, DC: The World Bank.

———. 2008. "Miss-Targeting or Miss-Measurement?" *Economics Letters* 100:9−12.

Ravallion, Martin, and Kalvin 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? Some Evidence for Rural India." In Dominique van de Walle and Kimberly Nead, eds., *Public Spending and the Poor: Theory and Evidence*. Baltimore, MD: Johns Hopkins University Press.

Stifel, David, and Harold Alderman. 2005. "Targeting at the Margin: The 'Glass of Milk' Subsidy Programme in Peru." *Journal of Development Studies* 41(5):839−64.

Smolensky, Eugene, Siobhand Reilly, and Erik Evenhouse. 1995. "Should Public Assistance be Targeted?" *Journal of Post Keynesian Economics* 18:3−28.

van de Walle, Dominique. 1998. "Targeting Revisited." *World Bank Research Observer*, 13(2): 231−48.

Weisbrod, Burton. 1970. "Collective Action and the Distribution of Income: A Conceptual Approach." In R.H. Haverman, and J. Margolis, eds., *Public Expenditure and Policy Analysis*. Chicago, Ill.: Markham.