Discussion: Targeting is Making Trade-offs

Jean-Pierre Habicht and Edward A. Frongillo

Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853-6301

ABSTRACT The previous articles presented different aspects of targeting: the implicit political implications, the trade-offs in giving power to different stakeholders to decide and to implement targeting, perceptions of frontline workers in implementing a program, and a technical article about selecting a scale for targeting, which we review in greater detail. It is well recognized that targeting results in a trade-off between not serving those who should be served and including those who should not be served. Less well recognized are the trade-offs that are the consequences of deciding between using indicators of risk vs. using indicators that predict benefit. J. Nutr. 135: 894–897, 2005.

KEY WORDS: • targeting • nutrition • programs • risk • benefit • trade-off

The articles issuing from the symposium on “Challenges in Targeting Nutrition Programs” were much broader in scope than we originally envisaged. They open new vistas of thinking and research. One broad vision about targeting (1) reveals that this apparently technical enterprise is enmeshed in political considerations, which are often hidden, but are amenable to analysis. Another broad vision (2) considers the trade-offs in involving different stakeholders in targeting decisions, where the political process plays out in tension with technical considerations of effectiveness and efficiency. Both these articles discuss the different ways targeting can be accomplished but from different complementary points of view. Finally, another article (3) described the perspectives of the frontline providers of a targeted national nutrition program.

One article (4) addressed the more technical issues that we had expected. It is a report about the Institute of Medicine’s (IOM) (5) examination of scales to measure “inadequate diets” had expected. It is a report about the Institute of Medicine’s (IOM) examination of scales to measure “inadequate diets”. The committee discusses the different ways targeting can be accomplished but from different complementary points of view. Finally, another article (3) described the perspectives of the frontline providers of a targeted national nutrition program.

The 2002 report itself from the IOM (5) gives more background. We then follow with a more detailed discussion about developing screening tools for targeting.

Pelletier (1) discusses how the premises underlying technical work are political. He uses a “liberal” and “conservative” dichotomy as an example. One can examine the motivations of the stakeholders involved in the WIC program, identify their interests, and then derive the approaches they would take to address the problem of dealing with the 2 “major concern” cells in Figure 4 of Pelletier’s article (1). The genesis of the WIC program involved many stakeholders, including those who were more interested in selling food than in social benefits. One may presume that those interested in selling food would favor more inclusive targeting, because more food would be sold to the government. Expanding WIC also corresponds to the interests of those responsible for the program if they follow a natural bureaucratic instinct, which is rewarded for expansion. These interests are compatible with those concerned about ensuring adequate health care and nutrition to all Americans, who put a higher premium on being sure that all who need WIC receive WIC, even if many who do not also receive WIC services.

The interests of those who favor expansion and more inclusiveness are in competition with those whose interest is less government expenditure, who thus favor a WIC program that costs less. One way of reducing costs is to diminish the number of beneficiaries through restrictive targeting. Similarly, technocrats who seek to improve efficiency would also impose a more restrictive screen, because that will increase the proportion of those served, who actually need the program.

The above review of players includes those who initiated WIC and those who fund and regulate it. The frontline workers who implement the program are usually omitted from these considerations. The article by Lee et al. (3) shows how important these players’ actions are in how a program is actually implemented. It appears from their work that, most often, frontline workers are concerned with ensuring that services are...
maximally accessible to cover those who are in need. The article by Lee et al. (3) about the Elderly Nutrition Program and the work by Dickin (6) on the Expanded Food and Nutrition Program, both administered by the U.S. Department of Agriculture, also illustrate well the wealth of understanding of the frontline workers and suggest that taking their knowledge about context, motivation, and practice into account in larger decision-making processes would improve the process and might well improve reaching the goals. This discussion illustrates that integrating the perspectives and the motivations of stakeholders requires historical, political, anthropological, and organizational psychology expertise. This is worth doing to bring these disparate views into a common framework for a discussion that includes not just the technical but also other perspectives, such as economic, legal, bureaucratic, and political. Our expertise, however, lies with the more technical aspects of developing a screen for targeting.

The first technical issue in targeting is the development of a valid screen to decide whom to include or exclude from an intervention or program. The screen can be developed to decide about where and when to intervene at population levels, such as in famine prevention (7), or it can be developed to intervene in targeting interventions at the individual level, such as preventing some deleterious behavior, such as poor eating pattern, or an outcome, such as death.

There are various approaches to building a reliable scale once the objectives and context are defined. One investigates the reliability of achieving some cutoff on a scale where both the cutoffs and the scales were constructed for other purposes. For instance, one might use dietary recommendations intended for counseling. Such recommendations are hortatory rather than normative in that many in a population may not follow the recommendations, and are unlikely to do so, and yet many of them will not suffer because of that lapse. The IOM 2002 report (5) illustrates the difficulty of adapting this counseling information for use in targeting a program. Another approach first builds the scale and then applies cutoff points. One such approach tries to measure exactly the underlying dichotomous construct (e.g., death or a long-term adequate diet) and then builds a feasible scale relative to that construct. Yet another approach is to develop an index through principle-component analyses, which uses a number of measures that seem to be related to an underlying but unmeasured construct and then examines whether they cluster into patterns that are consistent with the construct (8). This was the approach used in developing a scale for hunger (9,10).

Whatever the scale, it must be tested by the receiver operating characteristics (ROC) method against some “golden standard” of reality. This is a recognized method for dichotomous scales (11,12) such as the presence or the absence of a symptom, or a diagnosis. It is also the only method to be used for continuous scales (13), such as the scales usually used in nutrition targeting. We give an example with anthropometric data from 1-y-old Bangladeshi children measured in 1974 relative to their subsequent 2-y survival (14). The sensitivities and the specificities for identifying children who would die were calculated across the whole range of screening cutoffs for height-for-age, for weight-for-height, and for arm circumference (15). Figure 1 presents the ROC of those measurements by plotting Z-scores of the percentiles of sensitivity against the Z-scores of the percentiles of specificity in predicting the 2-y mortality. The interval scales of the Z-scores are presented in the left and bottom axes. The corresponding percentiles are presented in the upper and right axes to show that they are not plotted on an interval scale. The further the ROC line is from the indifference line, the better is the scale at predicting the underlying reality, in this case risk of death. The indifference line is the 45-degree diagonal going through the 50th percentile of both sensitivity and specificity. Figure 1 shows that stunting (height-for-age) and arm circumference were better measures for predicting death than was wasting (weight-for-height). Examination of the ROC lines plotted as the interval Z-scores is essential before proceeding with further statistical analyses, which are based on the assumption that the lines are parallel to each other and to the indifference line. In this example, the height-for-age and the arm circumference ROC lines are indeed parallel, but one cannot count on this usually being the case (13). The ROC lines for height-for-age and for arm circumference are not parallel to the indifference revealing that these 2 indicators are better screens at high sensitivity than at high specificity. The results from the ROC indicate that at higher sensitivity for height-for-age such as 80%, the specificity is 35%, and, at 80% specificity, the sensitivity is 48%, indicating a slightly better screen at lower than at higher specificity. This is a small deviation from parallelism and does not preclude using a single statistic (13) to describe the quality of the screen. In spite of being small, the deviation is still visible in Figure 1, which would not have been the case if the ROC had been plotted with an interval scale of percentiles. Such an inappropriate plot will miss ROC lines that are much less parallel when a single statistic makes no sense.

Using the above example, one can compare the proportions of deaths in those selected at cutoffs that select with 80% sensitivity and with 80% specificity, respectively. The proportions of deaths among those selected are called the “positive predictive value” by epidemiologists (16) and “yield” (17) by others. The yield depends not only on sensitivity and specificity but also on the incidence (or prevalence) of the outcome of interest in the population. In this case 112 of the 2019 children died for a population incidence of 55 deaths per 1000. At high specificity, the yield was 11.5 deaths per 1000 among those selected, whereas at high sensitivity, it was 55 per thousand, no different than for all the children. This example reveals how increasing specificity increases the yield. It also decreased the sensitivity to <40%, however, meaning that over two-thirds of the deaths were not predicted by this cutoff.
This example also reveals how poorly a screen may actually predict an outcome of interest even when an indicator scale (e.g., height for age) significantly correlated with the outcome, if the population incidence or prevalence is low.

The above example is also useful because it describes a high-specificity screen that identifies a group of children who are at higher than usual risk. It is true that some screens are designed to identify those who deserve help, such as the working poor, even if they cannot benefit from the program; however, in this case one would expect the screen to select those children who would benefit from the program by surviving. This requires a different reality than death; it requires that the reality be deaths prevented by the program. In other words, it requires identifying indicators that reveal a potential to benefit from the intervention (18). Thus, potential-to-benefit indicators are often different from indicators of risk, which reflect or predict future harm. For instance, mother’s height and head circumference are good predictors for risk of low birth weight but are poor predictors of benefit from nutritional intervention, because targeting on mother’s height will not improve the efficiency of nutritional interventions to prevent low birth weight (19). Similarly, height in infancy is a good predictor of height in adolescence in an undernourished population, but it predicted benefit in growth in height from food supplementation less well than did weight in infancy (20). These studies to identify predictors of benefit all required interventions.

In the absence of interventions studies to estimate potential-to-benefit indicators, one can estimate a potential (P) to benefit from the product of the risk (R) times the effectiveness of the intervention (E) to alleviate or to prevent the risk (i.e., = R*E) (17). This requires that the indicator of risk be appropriate for the intervention planned. Often the risk scale is a measure of some biological or behavioral reality, which is related to the risk. It is instructive that occasionally such a risk scale may be a poor measure of the underlying biological or behavioral reality, yet be excellent for predicting the risk itself. For instance, in our work to prevent fannies in Indonesia, we used expected harvest yield, because it was a good predictor of fannies. The validity and the reliability of the scale as it related to harvest yield suffered when the responsibility for collecting the information on harvest yield passed to those who were rewarded for the yield that they reported. They overreported unless the harvest was failing, in which case, they underreported, because they were rewarded for high yield and were punished for low yield unless the low yield was catastrophic. This invalid and unreliable indicator of harvest yield would be a better indicator of famine, because it included not only some knowledge of the real yield but added knowledge about other factors that were more pertinent to predicting famine than to harvest yield itself. Thus, a scale need only be valid and reliable at the cutoff point. This permits the development of scales that are neither continuous nor measure a same variable across the range of the scale. For instance, a cutoff on a Gutman scale (21) that uses the presence or the absence of a single attribute, such as a single household possession, or knowledge about a single fact, may be a good screen because that cutoff encapsulates the information necessary to target. This case of a biased estimate by frontline workers of a determinant of fannies, which is nevertheless accurate and reliable in predicting fannies, illustrates how frontline workers can sometimes target better than can quantitative indicators collected by other means.

In summary, the quality of the screen for targeting depends on the degree to which the screen concentrates those who can benefit from the program (1 minus specificity), with the least loss of those who could also benefit but are excluded from the program (1 minus sensitivity). One recommendation is to first identify the best scale for screening by ROC graphing and analyses, and to then decide on the cutoff point. This is difficult if alternate scales cross. The best scale will depend on whether one chooses the cutoff point above or below the crossover. Fortunately, the following considerations resolve the problem. The best cutoff point is one that results in including the number of people that the program can handle. A more inclusive and therefore sensitive screen will deliver more people than the program can handle at a particular time. A more sensitive screen will admit more people who are less likely to benefit from the program and who will displace those who are more likely to benefit if the number is exceeded. Thus, a screen that has too high a sensitivity for the current capacity of the program actually decreases the sensitivity of the program itself. This discussion reveals that using sensitivity and specificity as the basis for setting cutoffs for program screening is, in fact, incorrect if the number of actual beneficiaries is fixed. In that case, the basis should be the number to be delivered by the screen. Once the cutoffs are chosen, plotting them on the ROC lines, such as in Figure 1, will automatically identify the best technical scales. At this juncture, other characteristics of the scale, such as ease of measurement, can be taken into account in a final choice.

The above technical discussion holds when a program can only accommodate a specifiable fixed number of beneficiaries. In that case, there is actually no disagreement on where the cutoff point should be set between those concerned about including the most “needy” and those concerned with excluding the “non-needy.” This technical conclusion is very different from what one would infer from the “liberal–conservative” tension describe by Pelletier (1) once a program is established with a fixed number of participants. The numbers a program can handle, however, is not a purely technical issue, nor is it necessarily fixed. Sociopolitical processes determine the number. In that context, there is a real tension, because a more or less sensitive screen can be used to advocate or to oppose a program.

In this article, we have identified a well-known trade-off that is, in fact, not technically true if social-political considerations are absent. Thus, sociopolitical considerations entail trade-offs that, in turn, have implications for technical trade-offs. Many other trade-offs remain. At the technical level, decisions need to be made about using highly quantifiable measurements and using more holistic selections made by knowledgeable frontline staff. These decisions also need to take into account how much autonomy that staff should have in other aspects of the program. At a higher level of consideration is whether risk scales can substitute adequately for the potential to benefit scales, which are the scales that should be used but about which we are mostly ignorant. Developing better screens depends on developing potential-to-benefit scales, which is urgent, but presents major research challenges both relative to feasibility and, above all, to generalizability.

ACKNOWLEDGMENTS

We thank Gretel H. Pelto for discussions and insights, and Jennifer Schaub for creating the figure.

LITERATURE CITED