

## Session II: Methodological Issues in Food Security Measurement

### Testing the Robustness of the Food Security Scale With More Recent CPS Data

James C. Ohls

Abhijay Prakash, Larry Radbill, and Allen Schirm, my colleagues at Mathematica Policy Research, Inc., contributed to this work. Earlier Chris Hamilton described the Rasch model that provides the framework for the food security research that Abt Associates, Inc., did for FNS. A fundamental tenet of the model is that the underlying food security scale stays constant over time and that individual items stay more or less in the same place on the scale over time. It is important to test this underlying hypothesis to make sure it holds true in the current application. Under contract with FNS, we looked at whether the metric appears to be the same over time. Chris's early research was based on the 1995 data. Since multiple years of data are now available, our mandate was also to use the 1996 and 1997 data to expand the previous analysis of levels of food security to include a longer time period.

One technical issue raised by these research questions concerns screening. The Current Population Survey uses screening questions to track households into the detailed food security module; only something like 20 to 25 percent of households get tracked into the module each year. The screening questions were different each of the last 4 years, in part to experiment with different alternatives and in part to satisfy different constituencies. This raises the possibility that changes in results over time may be due not to some underlying phenomenon, but instead to what particular set of households enter the detailed analysis in any given year. To compare the 3 years, we identified what we call the "least restrictive common screen." This is defined as the least restrictive set of screening characteristics—that is, the set of characteristics allowing the most households to pass the screen—such that a given household will have the same screen-

ing outcome (pass and not pass) each year, if its values for the screening characteristics do not change. That is, a household with given characteristics that passes the original screen plus our screens in 1995 will also pass it in 1996 and 1997. The various screening criteria were nested in such a way that it was possible to develop this least common screen, to ensure that we dealt with completely comparable households in the analysis for all 3 years. The food security estimates we present here for 1995 are not quite the same as Chris's because we are using the least common screen and, therefore, a more limited set of households.

Another technical issue is normalization. Any linear transformation of a given Rasch scale has the same information content and yields the same results as the initial scale. A household's numerical score has meaning only relative to other numerical scores: such scoring attributes as the mean or the low-to-high range of the scores can be chosen by the investigator. Accordingly, to compare scores from different years, a single normalization must be chosen to ensure the same metric across years. The results we are showing today are based on setting the scale so that the mean of the item severity levels is zero. A second normalization used for much of the work follows the educational literature in setting the slope of what is known in the Rasch model literature as the "item characteristic curve" equal to one at its inflection point. That treatment comes close, at least in our data, to being equivalent to making the standard deviations of the item scores a constant. We are not using the zero-to-10 numerical range that Chris used for the scale.

To assess the effects of our screening on the model estimates, we compared our 1995 estimates made by using the least common screen with the Abt 1995 estimates. Our estimated ordering of items by severity is virtually the same ordering that Chris Hamilton and the Abt team obtained. We replicated their ordering with one exception: the ranking of two of the items was inverted. The use of different screening conven-

tions is one of several technical differences between this work and the work of the Abt team, but it is almost certainly the one that inverted the items. It is reassuring that in the Abt analysis those two items were clustered at almost exactly the same place on the scale, differing only at the second decimal place and by an amount that is not statistically significant.

We compared the 1996 and 1997 scores with the 1995 model to see whether the estimated ordering of the severity of each item stays the same over time. The basic result is that the ordering remains essentially constant. We continue to see the same inversion of two items as in the 1995 data, and then there is one other inversion that emerges for 1996 and 1997. The two items involved in this second inversion were so close in the original Abt analysis that their placement was almost indistinguishable; they just happen to be very close in the opposite direction in the later years.

In summary, the item order is essentially preserved across years. Differences are not statistically significant. Furthermore, the lesson of the results is that differences across years for any individual item are by and large not statistically significant. There are only two items where the differences over time are at all statistically significant. Our reading of the results is that the model is effectively the same in each of the 3 years.

In addition to the assumption of temporal stability, the Rasch model has an underlying assumption of stability across population subgroups. When the model is estimated on the population as a whole, the implicit assumption is that subgroups are behaving or reacting to these questions roughly the same. It is important to know if that fundamental tenet is true. We investigated the issue with several different groupings. The one I'll discuss is the ethnic grouping, because it is perhaps the one with the most intrinsic interest and it turned out to be the grouping with the most differences.

Among whites, there are three inversions of items, two of which are those noted earlier for the population as a whole. For African-Americans, there is a triplet cluster of inversions, that is, three consecutive items are rearranged, in addition to three pair-wise inversions. Hispanics have four pair-wise inversions. In general, inversions for the subgroups involve clusters of items similar to the inverted items in the national model. Overall, then, the results for the subgroups are a bit more complicated, but they are not dramatically different from the original Abt model. The model for any subgroup is recognizable as the same basic model: items are not shooting up and down in different ways.

In assessing these results, there is a question of magnitudes. What amount of item inversion alters the integrity of the model as a useful measurement tool for various applications? Is research that involves the food security scale jeopardized by the magnitude of the changes reported earlier? These questions are not susceptible to statistical tests but instead require research judgment. We solicited the judgments of Professor Benjamin Wright, an extensively published expert on Rasch models at the University of Chicago, and of Robert Mislevy, a senior scientist at Educational Testing Service (ETS). ETS is a national center of item response theory models, and it has the contract for the National Assessment of Educational Progress, which relies heavily on these models and for which Mislevy has done extensive work. Both experts indicated that, in their judgment, the above results showed more consistency than is usually present in Rasch model applications. In Bob's words, "this is about as good as it gets."

Finally, we calculated changes in food insecurity and hunger prevalences between 1995 and 1997. The essence of the results, which are preliminary, is that there was an increase in food security over the 1995 to 1997 period. The pattern is a little puzzling. There is hardly any increase from 1995 to 1996, and then food security rates increased

substantially in 1997. Mark Nord has just recently obtained 1998 data, and it is not yet clear whether the trend continued. For the time period under study, the most obvious candidate for improved food security is the booming economy, although perhaps policy changes could have also improved food security. Food insecurity has not ended, but in these data, it has been reduced somewhat.

## Discussion

### Edward A. Frongillo, Jr.

Jim Ohls and his team provided very well-written manuscripts.

Let us imagine that we have developed 18 items, and we simply count how many items a household answers affirmatively and then rank households by that count. Suppose the items are exchangeable, so that an affirmative on one is exchangeable with another. We could classify households based on whether less than three items are affirmed, three to seven items are affirmed, *et cetera*. The problem with this procedure is that those cut points for the classifications would be arbitrary, which is not very satisfying if we have a set of exchangeable items. We would feel better using a second procedure in which the items are not exchangeable but instead are ordered by “severity,” and we base cut points on our understanding of severity. Then severity would be measured not by just the number of affirmative responses but by knowing which items are affirmed and that some of them are indicative of greater severity.

The Abt team used such a method on the 1995 data, and Mathematica Policy Research has done additional work here. To check how well the method was working, we could tabulate the numbers of affirmative responses and see if the ordering we expected is actually shown in the data. Jim Ohls did not show this in his presentation, but in his paper, they had some clever ways to see whether the ordering was preserved. The manuscript discussed that the safest way to normalize across the surveys may be to recognize

that the same 18 items are used, and we expect them to perform the same.

Now notice to this point I have not used the term “Rasch model.” A Rasch model relates to some observed variables, in this case dichotomous variables, with some unobserved food security status of households and some unobserved severity of the item. So we have things we observed, which are the items, and then we have the notion that in the background households are more or less food insecure. We want to know about food insecurity, but we cannot measure it directly—at least not routinely. (We can measure it directly if we want to. Anne-Marie Hamelin has done this in her study in Quebec, but the method is intensive.)

According to the model, we assume the items differ in their severity but that item severity is unobserved; we cannot know just by looking at the item what its severity is. In fact, given our knowledge about food security, a good idea about severity can be obtained by looking at the food security items. We use a questionnaire to infer the household food security status, which we cannot see directly in a questionnaire, by observing these variables that actually get measured.

The model assumes that food security status is a characteristic of the household, and which exact items we use should not matter in our determination of that. An assumption that is symmetric is that measurement of item severity should not depend upon what households happen to be in our sample. We also assume that all items discriminate in the same way among the households. Once we account for item severity, then the items are in a sense exchangeable. If one accounts for food security status of household, then the households are exchangeable.

What is the value of the statistical model? First of all, it allows comparisons of the results from different sets of items. We do not, in fact, have the same items for all households; we have 18 items for households with children and only 10 items for households without children. A few people—remarkably, only about 3 percent—do not answer all the items. Furthermore, there

could be variations in wording from year to year. In the future, we might actually change the items. The Rasch model is very useful by allowing us to make comparisons, despite such problems and changes.

A second advantage of the model is that it can compare different groups of people based, for example, on location or race and ethnicity or household composition. I do have concerns, however, about subgroup comparisons.

The Rasch model is very good at comparing across different sets of items when all those items are supposed to be measuring the same thing. The extra items for households with children, however, have a degree of severity that does not exist in our measurement tool for other households. For example, in the 10-unit scale that was in the Abt report, there is a 2-unit difference between the most severe adult item and the most severe child item, which means we do not have any items at the most severe end for households

without children. These items are not missing at random; rather, the range of severity is truncated if there are no children.

The second concern is whether, theoretically, the Rasch model is a good tool for comparing across subgroups of people, that is, whether the model can tell us if the measurement tool is operating in the same way across different subgroups of people. In particular, the prevalence of food insecurity among the elderly may be underestimated. There is a need for fundamental research on groups other than households with children to provide an in-depth understanding of food security and a foundation for measurement. We do not know how to ask about the most severe food insecurity for households without children. We do not fully understand the importance of a food-use component of food security for the elderly, in addition to the components of food availability and access understood to comprise food security across all age groups.

## Assessing the Sensitivity and Specificity of an Abbreviated Food Security Scale

Stephen J. Blumberg

This work was jointly authored with Karil Bialostosky, William Hamilton, and Ronette Briefel.

Surveys that operate under time constraints or financial limitations are likely to cut back on the 18-item scale. For example, the Urban Institute's National Survey of America's Families limited itself to four items. My understanding is that the Census Bureau's Survey of Income and Program Participation tried to limit the number to seven. Choosing which items to retain, however, has been somewhat haphazard. We worked with the Welfare Reform, Nutrition and Data Needs Working Group to take a more systematic approach toward the design of an abbreviated scale that was sufficiently valid for general population surveys of food security. This scale was then validated by statistical comparison to the larger 18-item scale.

A few principles guided our selection of the items. First, we knew that less than 1 percent of the general population is insecure with severe hunger. For a reasonably accurate measure of that prevalence (say, less than 20 percent relative standard error), the sample size would need to be about 3,000 or so. Given that a survey with financial or time constraints is also likely to have a sample size constraint, we combined the moderate and severe hunger categories into one overall category. In our work, the two categories of food insecurity are insecure without hunger and insecure with hunger. Because we do not distinguish between the two most severe levels of food insecurity, the most severe items add little information; the six most severe items were dropped.

Second, any short form should be able to classify households with and without children. A scale with questions specifically about children is necessarily weaker when used to classify households without children. We, therefore, excluded the

four remaining child-focused items, leaving eight items.

Third, we excluded the first item in the scale because 80 percent of respondents who affirmed any questions affirmed this item. Finally, we retained the least severe item that clearly identified each food insecurity category. Given the remaining items and our feeling that six items were probably the minimum permissible length for this abbreviated measure, we were left with four possible six-item scales.

We compared classifications from the four possible short-form scales to the classifications determined by the 12-month, 18-item scale. Data for evaluating the six-item scales were collected as part of the April 1995 CPS. There were not many differences among the four scales. On average, they correctly identified the overall food insecurity category for 97.1 percent of the households. Population estimates of overall food insecurity were off by no more than 2 percentage points with all four scales. The particular six-item scale best at classifying households also had the least bias, and we concluded that it was the best set of items to use for an abbreviated scale.

Using categorizations from the 18-item scale as the standard, 97.7 percent of all households were put in the same category by the short form, given that we combined the two most severe categories into one.

When the prevalence of a condition is low, a scale with high specificity will usually correctly classify most people. Indeed, of the households who were food secure according to the full 18-item scale, 99.4 percent were still classified as secure by the short form.

The sensitivity of the short form was also quite good. Of those households classified by the 18-item scale as food insecure, either with or without hunger, 92 percent were classified as insecure by the short form. Of those households classified as insecure with hunger, 84.7 percent were correctly classified by the short form.

The short form correctly classifies a large proportion of households, in part because a large proportion of the households responded negatively to all of the items. But, when we examined just the subgroup households that affirmed at least one item, the sensitivity and specificity of the short form still continued to be strong.

We had excluded child-focused questions. As expected, the short form's sensitivity and specificity were, therefore, lower for households with children than households without children—in both the full sample and in the subgroup—but they were still quite acceptable. We tried other six-item subscales that included items that were child focused, and none provided a significantly better overall classification ability than the short form that I have been showing you.

The prevalences of overall food insecurity and food insecurity with hunger were under-estimated with this short form by just 0.3 percentage points. The overall bias of the short form was greater for households with children than for households without children.

The full 18-item scale is the gold standard that should be used if resources permit. But if resources do not permit 18 items and your research goals do permit the combining of the moderate and severe hunger categories, then we would recommend that the six-item short form be adopted as the standard. This will enable us to have a universal surveillance instrument and to make meaningful comparisons across surveys.

## Discussion

### Prasanta Pattanaik

The short form for assessing food insecurity and hunger in a household is a very useful tool that correctly classifies an overwhelming proportion of the households. It will be a helpful instrument when limited resources do not allow 18 items.

One conceptual point can be raised for the 18- and 6-point scales. After each household has been given a score on a particular scale, what do we do with these scores? One possible use is to

classify the household into some broad categories. This has been done. An alternative use can be to construct an index of food insecurity for the entire group of households. Such an index will be analogous to indices of income poverty in the literature that has developed following the 1976 paper of Amartya Sen, this year's Nobel Laureate in Economics.

In the literature on income poverty, economists use a benchmark level of income below that which a person is considered poor. Then economists consider to what extent a person falls short of this benchmark (a person who is at or above the benchmark is considered to have zero shortfall). The shortfalls of the different individuals are then aggregated in some way to arrive at an index of poverty for the entire group of individuals under consideration. Using this general method, the literature on the measurement of income poverty has come up with alternative measures of income poverty, usually on the basis of alternative sets of axioms that postulate properties that a poverty measure should satisfy.

In the context of food insecurity, we have a scale on the basis of which we can measure the extent to which a household falls short of the ideal of complete food security. I was wondering whether one could measure, for each household, the shortfall from this ideal of food security, and then aggregate the shortfalls of the different households to arrive at a single index of food insecurity for the entire group of households. The underlying intuitive approach has been developed rigorously in the mathematical literature on the measurement of poverty and has been widely used in practice by economists. I was wondering whether this approach could be used to construct an index of food insecurity as an alternative to using the household scores for classifying the households into broad categories. Of course, categorization is important. It captures one of the dimensions of the phenomenon of food insecurity. However, we can also capture other dimensions by following the approach that I outlined. If we want to follow the route I described, then we can probably use even the short form of six items.

Note that in the short form of six items, there are specific questions that, by themselves, allow us to discriminate between households. Some of the questions ask households how often a particular problem occurred: very often or occasionally. “Very often” indicates greater severity of the

problem than “occasionally.” Therefore, even for specific questions, we have some scope for finer measurement that we can use in applying, in this context, the overall methodology used by economists for the measurement of poverty and deprivation.

## Problems With Estimating the Prevalence of Child Hunger

Mark Nord

This work was performed jointly with Gary Bickel.

Concerns about children are salient in this interagency Food Security Measurement Project for legislative, programmatic, and public-perception reasons. Our paper proposes an improved method for estimating the prevalence of children's hunger and identifying households with hungry children.

Much has been said about the 18 questions as they relate to the four categories of food security. Some of the items are referenced at the household level, while others are referenced at the adult level. Eight items specifically ask about children in the household. These 18 items form a single scale.

The scale's severe hunger category is widely used as a proxy for households in which children are hungry, if there are children in the household. Much research concludes that households protect children from hunger until hunger reaches a severe level among adults, and only then the children start sharing in it.

Using the severe hunger category as a proxy for when children are hungry is effective if the items capture a unidimensional phenomenon. However, a second dimension could make that use problematic. Abt concluded correctly, I think, that the items are generally unidimensional but not perfectly so. The first dimension is severity. Once removed, the next factor in the raw data is, essentially, the extent to which households trade off adult hunger against children's hunger. The second factor creates concern that the overall 18-item measure may not optimally identify households in which there are hungry children. Even if only a small percentage of moderate-hunger households have children's hunger, the national prevalence of children's hunger could exceed the amount proxied by the severe hunger category by a large proportion

because there are so many moderate-hunger households.

We pulled out the eight child-referenced items, scaled them by themselves, and set a threshold. The location of the threshold with reference to the children's items was analogous to the threshold's location in the 18-item scale. We estimated the prevalence of children's hunger using the 8-item scale and compared the results with the estimated prevalence based on the 18-item scale. Cross-tabulation showed that we were not looking at quite the same households in the two estimates, and we examined household characteristics to understand the differences.

First let's look at the dimensionality issue. We submitted item residuals to principal components analysis. Because the correlation matrix pertains to residuals, the first principal component should be considered the second factor in the raw data; the first factor is severity as extracted by the non-linear Rasch method. We plotted the factor loadings of the items with severity of the item on the left scale. It is clear what the character of this factor is. It is the extent to which children in the household are protected from hunger at the expense of more severe adult hunger.

There are eight child items. The proportion of households with children that affirmed an item ranged from a high of 13.6 percent for "We relied on only a few kinds of low-cost foods to feed the children because we were running out of money to buy food" down to two-tenths of 1 percent for the most severe item: "Children did not eat for a whole day." The Rasch methodology scaled very consistently using these items alone or using the same items in the same households but adding the adult- and household-referenced items. That result is not surprising but it is always reassuring when procedures yield expected outcomes.

To get a prevalence estimate we had to establish an appropriate threshold. We examined item calibration and the household scores. We set the threshold between four and five affirmed items. Households classified as not quite having children's hunger would typically have affirmed these three items and that they cut the size of

children's meals, but would have denied the children were hungry. Those affirming five would also have affirmed: "children were hungry because we did not have enough money to buy food." The two items are very close so that, in fact, of those that affirmed four items, probably half also affirmed that children were hungry and denied that they cut the size of children's meals.

The threshold is exactly the same place as the severe-hunger threshold for adults on the household scale. Conceptually too it is the same place as the moderate hunger threshold relative to analogous adult items. To be classified in the moderate hunger category, the respondent must affirm three adult reduction-of-intake items. To be classified as having children's hunger, the respondent must affirm three items indicating reduction of quantity among children.

Our estimates of prevalences are based on 1995 data, the only ones in the public domain as of February 1999. We did replicate this entire analysis with the '98 data at the time when it was still unedited and basically everything in the '98 analysis was completely consistent with what I am reporting from the '95 data.

Among households with children, the current measure tells us that about 0.9 percent are in the severe hunger category at the household level. Therefore, we expect them to have children's hunger. Using only the child items, we find that about 1.1 percent of those households have children's hunger. Comparing the 0.9 and 1.1 percent figures might suggest that the difference between the two approaches is small. However, these percentage figures represent 332,000 households by the current measure and 425,000 households by our new estimates. If we want to focus on households with children's hunger, you could argue that the difference is enough to care about. The new estimates of 425,000 households represent a 29-percent increase over the current measure of only 332,000, and so the difference is proportionately large.

The survey contains the question: "In the last 12 months, were the children ever hungry but you just couldn't afford more food?" If that single

item is used individually to measure the presence of children's hunger, then 671,000 households would be registered—about double the level of the severe hunger category. We do not advocate using a single-item instead of a multiple-item scale. Nevertheless, the number that results from the single-item scale provides a face-validity check on where we put the threshold; certainly it would be hard to argue that we overcounted households with children's hunger.

We cross-tabulated households using the two approaches to investigate whether the 8-item measure is simply more sensitive, that is, it picks up the same households as the 18-item measure plus some additional households. It turns out that the groups are not concentric but overlapping. Of the households in the severe hunger category, 24 percent or 80,000 households are not classified as having children's hunger by the child hunger measure.

Finally, we compared the two subgroups. In some ways, I think, the comparison is the most interesting part of the paper because it identifies a plausible reason for the second dimension. The results of the two measures are not just random, that is, that some households just show up as having severe hunger and others with children's hunger. There is some logic to this second dimension.

We studied the difference between the two prevalence rates, subtracting the 18-item scale from the 8-item prevalence, for various demographic and economic categories. Here I report mainly the bivariate analysis, but multivariate analysis was done as well. Single-parent households had a positive difference. One might argue that single-parent households are less able to protect children against hunger at the expense of adult hunger because the household has only one adult. Households with more children are also less able to protect the children, resulting in a higher level of child hunger than is detected by the 18-item measure. The strongest single factor is the age of the oldest child. If the oldest child in the household is 15 to 17 years old, it is more likely that children are also sharing in that adult hunger. That is not surprising. Interestingly, for the

group in which the oldest child is 6 years old or younger, the prevalence of child hunger in those households is lower than you would expect, given the level of adult hunger.

No systematic differences appeared when comparing boys and girls, in contrast to what might occur in some other countries.

Income is a major factor that affects relative prevalence on the two scales. Even if a high-income household registers food insecurity or hunger, its experience is likely to be episodic and short term, and children will not be sharing in the hunger. But for low-income households, hunger is a long-term phenomenon during which it is more difficult to protect the children.

In the bivariate results, black households have a higher incidence of children's hunger than non-Hispanic white households, but the difference disappears in the multivariate framework in which the difference is accounted for by the income difference. A higher prevalence for Hispanic households is found in bivariate and multivariate results, a result that calls for future research.

I made a metro and nonmetro comparison because I am a rural sociologist. Rural children are better protected in households with the same level of adult hunger, and that result persists even in a multivariate framework for reasons I do not know.

In conclusion, USDA, NCHS, and other agencies in this interagency group need to consider the wisdom of supporting a second scale to estimate children's hunger, using the same survey instrument. Although an extra scale creates extra explanations and work, intuitively, I think, the extra scale is easier to explain. Perhaps we could then drop the severe hunger category from our household-level measure, which is hard to explain. Ultimately, the extra scale might be better at estimating the prevalence of children's hunger, for research purposes and for identifying which households have children's hunger.

## Discussion

### John Cook

When we were first working on the 1995 CPS data, we wanted to address many questions but did not have time. The question of how to measure the prevalence of children's hunger was certainly one of them. Mark and Gary provide considerable improvement in the severe end of the food security scale.

I fully support the creation and use of a separate child hunger scale for several reasons.

We know that children, especially young children, are in critical periods of growth and development. For them, nutrition and food security are even more important than for adults. In the post-industrial era, we sell our thoughtware, our brain power. Food insecurity and hunger may reduce children's human capital accumulation, and they will be severely impaired as adults. Future research should clarify the roles of under-nutrition, food insecurity, and hunger on academic achievement, and on other measures of human capital. In addition, children are probably a sentinel group with regard to food insecurity and hunger; they can serve as an indicator of problems likely to emerge in the rest of the population.

The technical portion of the 1995 reports contains a review of literature on physiological indicators for hunger. The physical sensation of hunger, the painful or uneasy sensation caused by a lack of food, manifests heterogeneously across persons, but it can be subjectively, reliably reported. There are physiological correlates, involving emptying of food and nutrients from the stomach and upper intestine, established in the physiological and clinical nutrition literature. Therefore, a key to measuring hunger is to identify conditions that result in below-normal food intake. Children and adults differ physiologically. For example, the liver—where energy is stored largely for immediate use—has a different size relative to overall body size so that children have to eat more often or become hungry more quickly.

For all these reasons, accurately measuring child hunger is very important.

Quality of diet is extremely important because everyone, at all income levels, should eat five servings of fruits and vegetables a day. Mark and Gary are developing a scale that we can use to address quality of diet.

Finally, child obesity is a major problem in the U.S. population. Bill Dietz has raised two principles or hypotheses that might be addressed in future research using the scale. First, to prevent children from feeling hunger, a family might rely on a few low-cost foods that are also high-fat

foods. Fat is a way to make the foods palatable, and low-cost foods besides beans and rice tend to be prepared with high-fat content. The second hypothesis involves weight cycling. Food insecurity may be periodic, occurring, say, in the last week of the month when food security is low. After a family gets its food stamps, or its pay, it eats fairly well for a while. Over time this eating cycle contributes to weight cycling, and Dietz observed that during the feast part of the cycle, children can gain more weight that becomes ever harder to lose. The difficulties of physical activity among low-income families, especially in metro areas, compounds those of weight-cycling. Childhood obesity may be a result.

## Luncheon Address:

### What Can Be Learned From Past Research on Measuring Poverty, Material Hardship, and Child Development Outcomes?

Susan E. Mayer

When I first heard about the efforts to develop a measure of food security, I was skeptical. But I have been impressed by the effort to conceptualize and measure food security. I was asked to reflect on lessons we can learn about the measure of food security from our experiences, with other measures of important social phenomena. Rather than focusing on technical issues, I will focus on how the food security measure is likely to be used and interpreted.

As measured, food security is mainly a measure of relative, not absolute, food insufficiency. This is in contrast with many other important measures of social phenomena, including the official poverty measure and measures of housing adequacy, which are at least intended to measure absolute deprivation. For example, the official poverty line is supposed to measure a constant level of purchasing power or a constant living standard over time. It is changed annually only for changes in prices, not changes in tastes or distance from the average living standard. Thus in principle, the United States could eradicate poverty by raising mean income (so long as inequality did not increase at the same time).

Some people think that we should have a relative measure of poverty, one that reflects changes in tastes and norms, because they think that both absolute and relative deprivation affect people. Others are content with the concept of absolute poverty, but they think that the way we measure absolute poverty is all wrong. And some people propose something in between: an absolute measure periodically updated to reflect changes in needs and tastes and spending patterns.

The food security measure differs from an absolute measure of this sort. It is relative in

three ways. First, it is relative in its intent. Food security specifically includes “an assured ability to acquire acceptable foods,” not just any foods, and in “socially acceptable ways,” not just any old way.

Second, many of the questions that make up the food security scale are about deviations from expected or normal food intake patterns. This means that the deprivation measured by the scale changes as normal or expected food intake patterns change. For example, a couple of the components depend on respondents’ ideas about what a balanced meal is, or their ideas about how much they ought to eat.

In fact, Americans’ ideas about what people ought to eat have changed a great deal over time. In her book, *American Living Standards*, Clare Brown tells us that in 1918 the typical breakfast consisted of two homemade muffins, biscuits, or pancakes; two slices of bread with butter; 6 ounces of milk; 6 to 12 ounces of coffee for adults; oatmeal or two eggs for adults; and bacon or sausage for the men but not for the women. If this were the prevailing idea of a normal breakfast today, many people could not afford it. I have no idea what most people think of as a normal breakfast today, and I do not know if expectations vary by income. I doubt that normal food-eating patterns correspond to what nutritionists recommend or even on what they consider minimally adequate diets, because so many people do not get all the recommended nutrients on a regular basis. But, without some idea of what a normal diet is, it is difficult to know whether deviations from normal are likely to be harmful.

In a recent *New York Times* article, a low-income mother laments that she could not buy her children Nike shoes and Gap clothes, that this made her depressed and she, therefore, stole money from her employer, which landed her in jail. Her kids then had to go live with their grandmother. I do not mean to malign this particular mother. I am sure her circumstances were more complex than the article suggests. But few Americans would have much sympathy for someone because she could not buy their children Nikes or clothes at the Gap, even if that is normal in some places.

The fact that food insecurity emphasizes deviations from normal eating patterns also means that a homemaker who manages to consistently provide low-quality meals so she can avoid cutting back or have her family go without food may appear to provide more food security than a less competent homemaker who spends more on food but doesn't make it last, even when their families' food intake is identical. Because food security depends to a large extent on deviations from normal food intake patterns and not on the nutritional value of the overall diet, families with more inconsistent lifestyles are likely to have more food insecurity.

Third, the food security measure is relative because the seriousness of food insecurity depends on the prevailing food-intake patterns. When most families consume a low level of nutrients and calories, cutting back on food consumption is quite serious. When most families have an abundance of food and overeating is the most serious nutrition problem, cutting back on food consumption is a less serious problem. As countries get richer and normally eat higher quality diets, the deviations from normal become less severe, even if the deviations do not change in frequency.

According to Clare Brown's 1935 book, among low-income groups, 90 percent consumed too little calcium, over 80 percent consumed too little iron, over 80 percent consumed too little vitamin A, over 60 percent consumed too little vitamin B-1, and 75 percent consumed too little vitamin C. These and other deficiencies arose not mainly because families skipped meals or went a day without eating, but because their overall diets were woefully inadequate.

Today, as in 1935, there are important differences in the degree of food security between rich and poor Americans. One of the background papers alluded to this morning used the one-question measure of food insecurity to classify households as food insufficient or food sufficient. Food-insufficient households consumed 20 percent less vitamin C, 20 percent less iron, 12 percent less phosphorus, and 15 percent less thiamin than food-sufficient households. But both groups con-

sumed over 100 percent of these nutrients. The food-insufficient group also got less vitamin E, B-6, magnesium, and zinc than the food-sufficient group. But neither group got 100 percent of these nutrients. The absolute intake of the nutrients matters as much as the difference between the groups. The normal diet for food-insecure people today is probably superior to the normal diet of even food-secure people in 1935.

Other questions that are part of the food security scale have this same quality. For example, the seriousness of a positive response to "Did you ever eat less than you should because there was not enough money for food?" depends on the steady-state diet, which changes as countries get richer or as norms about adequate diets change.

I want to now turn to how people are likely to understand the measure of food security. Constructed measures that tap a concept, such as food security, for which most people have no clear intuition, can often take on a peculiar meaning. Constructed measures of concepts for which people do have a clear intuition also can be misinterpreted if the measure does not correspond with people's intuition. I think that is the case with, for example, measures of price changes. Everyone knows what a price change is—inflation is when things get more expensive. Yet measuring the exact amount of inflation is very difficult. The Consumer Price Index does not measure the everyday understanding of a price increase, yet that is what most people think it measures.

In classes, when I ask my students what they think to be poor means, they overstate the amount of material deprivation associated with poverty. They also tend to think of poverty as static—that is, as the same people being poor year after year. They tend to have two visions of poverty. One is the poverty in high-rise public housing and the second is rural shacks. These preconceptions about poverty, not its true nature, influence the political debate over what to do about the poor. Similarly, the popular view of food security or hunger, not the careful and narrow meaning that USDA gives it, will influence

the political debate over what to do about food security.

USDA has done, I think, a careful job of saying exactly what it means by food insecurity and hunger. And as long as those who use the measure are careful researchers and policymakers, the scope for misunderstanding seems modest. But the measure is bound to be used by more than the few who understand it.

The food security measure actually combines one concept for which people have a lot of intuition, hunger, and one for which they have little intuition, food security. I wondered whether the way the USDA defines hunger corresponds with how Americans will interpret the measure of hunger. So I chose a sentence from Andrews, Bickel, and Carlson's article in the *Family Economics and Nutrition Review*,<sup>7</sup> which I thought was carefully worded to convey a precise message and exactly the kind of message one might put in a press release. The statement is "There were 4.16 million households in which one or more person experienced some form of hunger in the 12-month period preceding April 1995."

First, I read this statement to eight people from the University of Chicago. This is admittedly a very small sample. I then asked each person what he or she thought hunger meant. All eight agreed that hunger meant that a person could not afford to buy food for some period of time. But they disagreed about what that length of time was. Most thought it meant that to be hungry a person had to go without food for more than a day because, as one person put it, "Going without food for a day won't hurt many Americans."

Some thought to be hungry, a person had to go without food for at least a day several times over some period like a year. Only one person thought there had to be some physical harm related to not eating for someone to be hungry. All in

<sup>7</sup> Andrews, Margaret, Gary Bickel, and Steven Carlson. "Household Food Security in the United States in 1995: Results from the Food Security Measurement Project," *Family Economics and Nutrition Review*. Vol. 11. pp. 17-28. 1998.

all, this seemed to me pretty consistent with the spirit of the USDA meaning of hunger.

Next I asked a different set of people not what they thought hunger meant, but what they thought the statement meant. These four replies capture the spirit of the responses:

"I do not get what you mean by a household being hungry. Aren't people hungry?"

"There are at least four million hungry people in the United States. No, even more than that because these are households."

"Does that mean that 4 million people were hungry for a day or 4 million on some day?"

"Well, it means just what it says. Four million Americans are hungry on any day . . . . Can that be right?"

Well, of course, it is not true and I will come back to that.

I also wanted to see how people would report a statement like this if they had, say, read it in the *New York Times*, then went to work and told a colleague who told another colleague. In other words, I wanted to see how this kind of statement would get translated in the conversations of people. I read the statement to two people, asked them to tell another person what I had said, then to ask that person to tell yet another person, who would then come tell me what they had heard. Thus, I got two responses. The first was "One person in 4 million is hungry on any day in the United States." The second was "On any day in 1994, over 4 million Americans went hungry."

The way food security is measured does not allow a very precise estimate of how many people are hungry on a day. But we can estimate that if there are 4.16 million households with two people who are hungry for a total of 3 days each per year (for a total of 6 hunger days), this averages about 70,000 people who are hungry on any 1 day. The actual number could be double that or half that, but I am pretty sure that 4.16 million people are not hungry on any day.

The lesson is that the more you can express a measure in terms of how people actually think, the more likely they are to understand it. People think of hunger as an attribute of individuals, not households. If you said in 1995, X million parents reported that their child went a whole day without eating because the family couldn't afford food, almost everyone would understand what it meant. The idea that someone was hungry some time over some period is not very intuitive to people, so they are likely to misinterpret a measure of that concept.

Thus, there could be a public relations problem in the way the food security measure is used that might cause misunderstandings about the amount of hunger and food insecurity in the United States.

Not only is the measure of food security likely to be misinterpreted, it is also likely to be misused just as the poverty rate and other measures have been misused. A couple of potential misuses make me nervous. The first is using the food security measure as though it were a more general measure of economic distress. I have already seen a couple of papers that do this. It is very important to be clear that a measure of food security tells you only about food security. It does not tell you about overall economic distress or material hardship. Imagine two families who are equally well off. One skips meals to be able to pay the rent. The other fails to pay the rent so that all the family members get all their meals. If we only looked at food insecurity, we might think that the first family was worse off than the second.

The point is that if we want to know how many people are hungry, we cannot infer it from how many are poor. That is exactly why so much effort went into developing this measure of food security. But it is also the case that if we want to know how many are poor or economically distressed, we cannot infer it from the food security measure.

If we really want to measure economic well-being, material well-being, or living standards, we would need to put our minds to doing just that.

Using the food security measure to assess progress in the Food Stamp Program is also likely to cause problems. First, it is hard to imagine that food stamps will further reduce the overall incidence of food insecurity. It appears that only about half of the households reporting food insecurity are close enough to the poverty line to get food stamps. And many households below the poverty line who report food insecurity are already getting food stamps. Furthermore, food stamps are basically an income transfer and many of the causes of food insecurity seem to be related to things other than income. For example, holding poverty status constant, food insecurity declines with age, is lower for Asians and Pacific Islanders than for other races, and is greater for families with children than for families without children. This implies that learning to manage a budget and to prepare food, having lots of time, and having some types of food preferences rather than others are related to food security. Food stamps can hardly be expected to change these factors.

Finally, it is not clear that, in the current political climate, success in reducing food insecurity with government programs will be viewed as success at all. The definition of appetite in the *Devil's Dictionary* is "Appetite is an instinct thoughtfully implanted by providence as a solution to the labor problem." This definition seems to correspond to current views.

This brings me to my final point and that is that politics matters. Virtually no one thinks that the official poverty line is right, and there is considerable consensus about how it could be changed for the better. Yet no changes are on the horizon. The Consumer Price Index almost surely overstates changes in prices. The technical issues associated with measuring inflation are complicated, but still changes are slow to come. The reluctance to change these measures comes from the fact that both have great political prominence. This prominence is a sign of their success; unsuccessful measures do not get political attention.

By all indications, the food security measure is already becoming successful, at least in the sense

that it gets a lot of attention. It will, no doubt, also become politicized. Once it is, it will have a life of its own. If you think people at the University of Chicago misinterpreted the hunger-prevalence sentence that I read you, that is nothing compared with how it will be misinterpreted on the floor of Congress. The measure will end up misused and abused. No one will like it and no one will want to change it. Once politicized, all the careful planning and framing of the idea of food security will be lost. Politicians and advocates will change the meanings of food security to serve their agendas. Some will claim it understates hunger and food insecurity. Others

will swear it counts too many. Half will scoff at the measure without having any idea how it is created. The other half will use it as though it had no limitations. Academics will find every flaw. Meetings will be held, conferences convened, task forces organized, and recommendations made. And the measure will endure. No critique will be enough to get it changed.

Now this is perhaps yet another way that appetite resolves the labor problem. It means that we are all secure in our jobs. It also means that we will have yet another occasion to meet for lunch, I am sure.