Methodological Findings and Early Conclusions Based on the 1995, 1996, and 1997 Food Security Data

James Ohls, Abhijay Prakash, Larry Radbill, and Allen Schirm

The publication of the U.S. Department of Agriculture's 1997 report on food security levels in the United States (Hamilton et al., 1997a) has spurred widespread interest in measuring food security for various groups in the U.S. population. Using data from the April 1995 Current Population Survey (CPS), that report presented a comprehensive method for measuring food security levels. Other major surveys that have measured food security or plan to do so include the Panel Study of Income Dynamics and the National Health and Nutrition Examination Survey.

The 1997 USDA report was based on a single CPS sample for April 1995. An important next step in food security research is to extend that analysis to later years and develop a method for measuring changes in food security over time. Important research questions include:

- Are estimated model parameters stable over time?
- How is the prevalence of food insecurity in the U.S. population changing?
- How robust are prevalence estimates to alternative ways of implementing the procedures used in the 1997 report?

A Brief Summary of the Literature Informing the Food Security Concept

Although hunger has long been a concern of American nutrition policy, attempts to measure it systematically have posed major challenges to advocates and policy analysts. Early attempts to equate hunger directly to malnutrition were not successful, because they encountered conceptual difficulties in defining malnutrition and operational difficulties in developing reliable and inexpensive ways of measuring people's nutrient intake. Furthermore, as additional discussion took place, it was recognized that feeling physical hunger, a sensation experienced by most people fairly frequently, is not equivalent to the social problem of hunger, a situation related to economic deprivation. Further development of the concept was needed.

From the late 1970s through the early 1980s, there was growing interest in broadening the concept of hunger to the more general construct of resource-constrained food insecurity. This broader concept came to be defined in terms of phenomena and experiences associated with being at risk of hunger as well as actually experiencing hunger. Lacking access to food because of resource constraints also came to be included in the consensus definition of hunger as a policy issue.

The broadening of the relevant concepts took place partly within the government, with the inclusion of sets of questions related to food insecurity in the two most recent administrations of the Nationwide Food Consumption Survey. Two private research efforts also gave substantial impetus to the evolving focus on food insecurity. First, the Community Childhood Hunger Identification Project (CCHIP), organized by the Food Research and Action Center and funded by local and national business and philanthropic organizations, demonstrated that reasonable and consistent answers could be obtained, using a set of survey questions designed to measure food insecurity (Wehler, Scott, and Anderson, 1995). Second, work at Cornell University provided additional theoretical support and advanced the development of measurement scales based on answers to survey questions about food security (Radimer et al., 1992).

Beginning in 1992, staff of the Office of Analysis and Evaluation within the USDA Food and Nutrition Service (FNS) began a systematic effort to develop a battery of questions about food insecurity that could be administered regularly in government-conducted surveys. Drawing on previous research findings about food insecurity, together with additional research commissioned from outside researchers, USDA staff assembled the full range of food security survey questions that were used and identified sets of items that had promise as reliable indicators. FNS was assisted in this work by an expert panel that included many leading food security researchers.

FNS passed an important milestone when it won approval from the U.S. Office of Management and Budget for a supplement to the April 1995 CPS containing a set of questions designed to measure food security. The supplement gathered information about households' shopping patterns and various aspects of food insufficiency and insecurity during the 30 days and 12 months prior to the interview.

In 1995, Abt Associates, assisted by staff from Tufts and Cornell Universities, was engaged by USDA to analyze the 1995 CPS data. Faced with a questionnaire containing more than 50 items, the Abt team worked with the USDA to further refine the underlying concepts of food insufficiency and food insecurity. Along with this conceptual work, the team had to identify which of the CPS questionnaire items measured food insecurity. In the early stages of their work, they relied heavily on factor analysis to identify a group of items that, taken together, appeared to measure food insecurity. Then the Abt team applied a scaling procedure (described later in this paper) to assign a food security measure to each household. Based on these measures, households were classified into four categories—food secure, food insecure without hunger, food insecure with moderate hunger, and food insecure with severe hunger—and the Abt team estimated the prevalence of these four levels of food insecurity.

This paper extends the research of the CCHIP, Cornell, FNS, and Abt researchers. The work of the Abt team was focused on developing and implementing a measure of food insecurity, using data from the April 1995 CPS. We now have data from two additional rounds of surveys: the September 1996 and April 1997 CPS.[1] With these additional data, the focus has begun to shift to issues that arise in the development of a major ongoing social indicator. Although some issues examined here might not be important if the prevalence of food insecurity were going to be measured only once, or once in a great while, they can be critical when prevalence is measured on a routine basis and changes in prevalence are being closely monitored by policymakers. The recent availability of food insecurity data from two additional years allows us to address issues that arise when tracking changes over time. We present our preliminary empirical findings after briefly discussing the data used in our analysis and the Rasch model used to assign food security scores to households.

The CPS Data on Food Security

The data for the current study come from the Current Population Survey (CPS). The CPS is a monthly survey of about 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics. The sample is designed to represent the civilian noninstitutional population. Each monthly sample is divided into eight representative subsamples or rotation groups. A given rotation group is interviewed for a total of 8 months: it is in the sample for 4 consecutive months, leaves the sample during the following 8 months, and then returns for another 4 consecutive months. In each monthly sample, one of the eight rotation groups is in the first month, another rotation group is in the second month, and so on.[2] Under this system, 75 percent of the sample was common from month to month and 50 percent from year to year for the same month.

[1]The food security supplement was also fielded in subsequent years.

[2]More formally, the CPS sample is actually one of geographic addresses rather than households. If sample members move to a new address, they are not interviewed at that new address and they thus leave the sample. However, the address those sample members moved from remains in the sample, and the new residents are interviewed. These are known as replacement households.

The primary purpose of the CPS is to provide information about the labor force characteristics of the U.S. population. In each month, however, a supplement is added to the core questionnaire. In March of each year, for instance, the U.S. Bureau of the Census sponsors the Annual Demographic Supplement. This survey is the data source for the official income and poverty statistics published by the U.S. Bureau of the Census each year. In April 1995, September 1996, and April 1997, a special supplement was added to the CPS core questionnaire that included questions about household food sufficiency, food security, food expenditures, and a number of other related items. The structure of the food security supplement used in these surveys was as follows:

(1) Food expenditures during the prior week;

(2) Participation in food assistance programs (food stamps, elderly meal programs, school meal programs, and the Special Supplemental Nutrition Program for Women, Infants, and Children);

(3) Food insufficiency during the prior 12 months and ways of coping with that insufficiency; and

(4) Food security during the prior 12 months and the prior 30 days.

Not all households were asked the full set of questions in the supplement. To minimize respondent burden, a set of preliminary screening questions was used to determine whether there was evidence that a household might have experienced any food insecurity. If there was no such evidence, most of the subsequent food security questions were skipped. Across the three CPS samples, different screening procedures were used.

Table 1 shows unweighted sample sizes for the three CPS samples used here.[3] The initial sample size for the April 1995 CPS was 53,665 households. Budget cuts in January 1996 resulted in reduced sample sizes. This is reflected in the initial sample sizes for September 1996 and April 1997 shown in table 1. The initial sample for September 1996 was 47,795, and for April 1997, it was 47,306. In all three samples, roughly 85 percent of the core households entered the food security supplement.[4] Of those households, about 40 percent of the April 1995 sample passed the screening questions and were asked the balance of the food security supplement. For the two more recent surveys, tighter screening procedures resulted in only about 26 percent of households being asked the balance of the food security supplement. In all cases, there is a presumption that households failing to pass the screen are food secure. The differences in the screening procedures used across these three samples have important implications for the consistent measurement of food insecurity, a necessary prerequisite for measuring change in the prevalence of food insecurity. These issues are discussed in more detail below. Most of the research reported below is based on 18 key questions (items) that are used for the measurement of household food insecurity. Households with one or more children are asked all of these questions; childless households are asked only the 10 items that do not pertain to children. Once the differences between households with and without children are taken into account, there was very little item nonresponse. In all three samples, more than 97 percent of the households that passed the initial screen responded to all of the items used to measure food insecurity that they were asked. Even so, the fact that childless households responded

[3]Tables are at the end of this paper.

[4]The sample attrition at this stage is due mainly to households in the CPS being told that they are about to start a new module and their declining to do so.

to only 10 of the 18 items asked of households with children presents an additional complication, as discussed below.

The One-Parameter Logistic Item Response Theory (Rasch) Model

The Food Security Supplement to the April 1995 CPS contained more than 50 questions. Of those, about six were used as a preliminary screen to identify households that showed no indication of any food insecurity during the prior 12 months and, therefore, were not to be burdened with additional questions. Of the remaining items, 18 are used directly to measure households' food insecurity levels over the prior 12 months. (Of the questions not used, some apply only to the 30 days prior to the interviews and others were found during preliminary analysis not to be useful in developing the full food insecurity scale.)

In the first round of research on the 1995 data, it was desired to develop a method for combining answers on the 18 items into a single scale measuring household food security. In doing this, it was necessary to take the following factors into account:

(1) Not all questions applied to all households; in particular, 10 of the questions were not relevant to households that did not have children; and

(2) The data included some item nonresponse, involving households that did not answer all questions relevant to them.

In developing the desired food insecurity scale, the researchers involved drew heavily on a rich body of procedures used in the educational testing literature called Rasch modeling and item response theory (IRT). IRT methods have been widely used in educational contexts, such as the Scholastic Aptitude Tests (SAT) and the National Assessment of Educational Progress, to measure student attributes (such as math ability), using tests that, for test security reasons and other factors, are not identical.[5] In applying similar methods to the food insecurity measurement context discussed in this paper, the attribute being examined is food insecurity, and the test items are the individual food security questions on the CPS supplement.

The methods that are in the original analysis of the 1995 food security data and that are applied in the current paper involve a closely related technique called Rasch modeling.[6] The salient characteristic of the Rasch model is that the model involves estimating only a single parameter, often called the severity level, with which to characterize each question on the scale. Other versions of IRT theory estimate either two or three parameters per question.

The appendix provides a more detailed summary of the Rasch model. We conclude this section by noting certain salient properties of Rasch models that are relevant to the discussion below:

- The scale measure for households with complete data can be calculated based only on the number of questions about food insecurity that they answer affirmatively.

- The scale measure determined by the model are only unique up to a linear transformation; once a scale is developed, any linear transformation of the scale conveys the same information.

- In a Rasch model, each household's level of food security and each item's level of severity are items to be estimated in the model.

[5]For summaries of IRT theory, see Hambleton and Jones (1993) and Wright and Masters (1982).

[6]Some researchers view Rasch modeling as a subset of the IRT theory; others disagree with that characterization. In any event, they are clearly closely related.

Methodological Issues

Several important issues arise in constructing food security estimates from the 1995-97 data. These include the following:

- Screening households for evidence of food insecurity.
- Treatment of households with missing data.
- Estimating the standard errors of the estimated parameters.
- Whether to use weighted data in the estimation work.
- Standardizing the scale into a common metric across years.
- Establishing cutpoints for classifying households into various food insecurity categories.
- Choice of software to use in estimating the Rasch model.

This section discusses these issues and presents preliminary recommendations and empirical findings pertaining to them.

Screening Households for Evidence of Food Insecurity

As noted previously, to reduce respondent burden, the survey asked several screening questions to determine whether households should be asked the full battery of food security questions. Households that failed the screen—i.e., showed no significant evidence of food insecurity in their screening question responses—were skipped past the subsequent detailed questions and were classified as food secure in the subsequent analysis.

The screening questions used were different for each of the 3 years. If no adjustment were made for these differences, there would be the risk that similar households would be treated differently in different years. This could happen, for instance, if a household that in one year failed the screen and was classified as food secure would in a different year have passed the screen and possibly been found to be food insecure, based on answers to subsequent questions. This lack of consistency across years could confound attempts to examine changes over time in rates of food insecurity.

To avoid this problem, for much of the analysis in this paper, we have applied common screens for the 3 years. These common screens are based on questions that are used in the screens for all 3 years, and they ensure that the households passing the screen have all provided consistent answers to the same set of questions. Two possible ways of specifying common screens are considered. We discuss these after describing the individual screening questions used.

The Screening Questions Used

In all of the relevant years, the following screening questions were used in various combinations:[7]

- Did [the household] sometimes or often not have enough to eat?
- Did [the household] run out of food in the previous 12 months?
- Did [the household] feel that it sometimes did not have the kinds of food it wanted to eat?
- Did [the household] ever run short of money and try to make its food or money go further?

We have examined the screening questions in detail to identify the "loosest common screen," which we define as the screen that: (1) can be commonly applied to all households in all 3 years, and (2) allows the maximum number of

[7]The discussion in the text paraphrases the relevant questions. The exact wording of the questions can be found by accessing the CPS Food Security Supplement questionnaires at: www.ers.usda.gov/data/foodsecurity/cps/index.htm

households to pass the screen.[8] This loosest common screen is summarized in the Loosest Common Screen box.

As can be seen in this box, the only difference in the treatment of low- and high-income households is that the high-income households have to meet both—rather than just one—of the last two criteria, to pass the screen on the basis of these factors.

The loosest common screen allows maximum use in the analysis of households' responses to food security questions, subject to the constraint of commonality of screening criteria across years. However, the lack of symmetry between low- and high-income households may be troubling to some. Furthermore, the implementation with these data of screening criteria based on income is problematic, due to inadequacies in the income information available on the CPS. In particular, the income data are based on information supplied by the households when they entered or rotated back into the CPS sample, which could have been as much as 3 months earlier. Two other problems are: (1) that all types of income are included in the same question, and (2) that households are asked to respond only in terms of broad income intervals, e.g., $10,000 to $12,499.[9] These factors mean that it is likely that there may be substantial errors in the data. Another problem is that the interval end points are not changed from year to year. With inflation, there is drift over time in the real values of the interval boundaries, leading to differences across years in the real incomes that distinguish low- and high-income households.

Another potential drawback of using the loosest common screen—a drawback that is only applicable to low-income households—is that some analysts believe that the third criterion listed in the upper part of the Loosest Common Screen box ("didn't have the kinds of food wanted") may, by itself, be too ambiguous to be an appropriate basis for allowing households to be classified based on the full battery of questions. There are many possible reasons, other than resource constraints that prevent effective access to food, for not having the kinds of food a household wants. For instance, household members may be on diets to lose weight or may have unrealistic standards about what constitutes a good diet. As a result, it may be better to couple the "didn't have the kinds of food wanted" criterion with the "tried to make food or money go further" criterion, as is done in the Loosest Common Screen box for high-income households, so as to emphasize the resource-constrained aspect of food insecurity.

Loosest Common Screen

Low-income households

- Sometimes or often not enough to eat;
- Ran out of food in last 12 months;
- Didn't have the kinds of food wanted; or
- Tried to make food or money go further.

High-income households

- Sometimes or often not enough to eat;
- Ran out of food in last 12 months; or
- Didn't have the kinds of food wanted; and tried to make food or money go further.

[8]The first condition means that the actual screen used in a year was never tighter than the common screen. In other words, a household could not pass the common screen and fail the actual screen.

[9]Once a year, in March, the CPS collects detailed income data. However, none of the food security supplements have been administered in March. Although data from different months can be merged for some households, doing so involves substantial loss of sample, because there are households rotating into and out of the CPS sample each month. The loss is 25 percent for the 1995 and 1997 samples, and 100 percent for the 1996 sample due to the rotation group design.

A different potential set of screening criteria, which is intended to avoid these difficulties is summarized in the Tighter Common Screen box.

This tighter common screen applies to both low- and high-income households the criteria applied to high-income households by the loosest common screen. This tighter screen thus completely avoids reliance on the CPS income data and is also less reliant on the "didn't have the kinds of foods wanted" criterion.

Table 2 shows the number of households that pass the alternative screens and thus have their food security status determined using their answers to the full battery of food security questions, rather than being classified as food secure based on the screening questions. If no common screen is applied at all, the number of households that pass the actual screen used—the maximum sample screen—is over 18,000 in 1995 and about 11,000 in 1996 and 1997. About half of the low-income households passed the loosest common screen in 1995, and all of them passed in 1996 and 1997. Whereas all of the high-income households passed that screen in 1995, between one-half and three-quarters passed in 1996 and 1997. All high-income households that passed the loosest common screen also passed the tighter screen because the screens are the same for those households. However, about one-third of the low-income households passing the loosest common screen fail the tighter screen.

Tighter Common Screen

Low-income households

- Sometimes or often not enough to eat;
- Ran out of food in last 12 months; or
- Didn't have the kinds of food wanted; and tried to make food or money go further.

High-income households

(Same as for low-income households above.)

We recommend using the tighter common screen, in part because of our concern about the quality of the available income data needed for the looser screen. However, a final decision has not been made about which screen to use. We examine below the sensitivity of the results to the choice of screen. Except in that sensitivity analysis, the results presented were obtained using the tighter common screen.

Treatment of Households With Missing Data

The rates of item nonresponse among the households tracked into the detailed set of food security questions is quite low. Fewer than 3 percent of respondent households failed to answer one or more questions that they were asked. Nevertheless, while item nonresponse is relatively low, decisions must still be made as to how to deal with it in analyzing the data. Alternatives are discussed below.

A convenient feature of the Rasch model is that it is capable of assigning household scale levels to households with only partial data. Essentially, it determines the best fit for a household, given whatever data are available. The results of this fitting process depend on: (1) the responses given to the answered questions, and (2) items with the missing data. The previous study based on the 1995 CPS included all cases in the modeling that had nonmissing data for at least half of the relevant items, and this approach was feasible and yielded reasonable conclusions.

However, a drawback of including cases with any missing items is that doing so significantly complicates the interpretation and analysis of the modeling results. The main reason for this is that including cases with missing data greatly increases the number of possible levels of estimated household food insecurity observed in a data set. In particular, as noted earlier, it is a property of Rasch models that for households with complete data, a household's estimated attribute score is based only on its count of affirmative answers. It follows from this property that, with a survey containing 18 questions, all households with nonmissing data will be classified into 1 of only 19 food security levels.[10] This is often very convenient in examining model results and in parsimoniously presenting analysis findings. However, once the possibility of missing data on individual survey items is allowed, and recognizing that there are many possible permutations of which items are missing, the number of possible food security measure values greatly increases, and intuition can suffer.

On the other hand, excluding observations essentially involves discarding data that can improve the accuracy of the model estimation work. On balance, our preliminary recommendation is to include in the analysis all or most of the cases with missing data; however, this is still subject to review and discussion. In the meantime, in the sections below, we examine the sensitivity of key results to alternative assumptions.

Standard Errors of Estimates

It is important to calculate standard errors of model parameters and prevalence estimates to assess their precision and judge whether estimated changes over time are statistically significant. We have developed approximations to the relevant standard errors and are continuing work to improve those estimates. The following discussion focuses on two sets of standard errors—those for estimated item severity levels and those for prevalence estimates.

Standard Errors of Item Severity Estimates

The Rasch model calculates for each survey question (item) an estimate of its severity, in terms of the level of the attribute (in this case, food insecurity) being examined. The available software programs for estimating the parameters of Rasch models calculate standard errors for each of these item severity levels. However, these calculations essentially assume that the data represent a simple random sample and do not take into account the complex design of the CPS. Therefore, it is likely that they underestimate the true sampling variability, since they ignore the effects of clustering of households in the CPS. We have developed software to use replication methods to estimate standard errors that account for the effects of the CPS's complex design.

Prevalence Estimates

The estimates of food insecurity prevalence rates presented in this paper are based on tabulations of the estimated food insecurity levels of the households in the CPS samples. There are two potential sources through which sampling variability affects these estimates: (1) variability due to sampling error in estimating the model parameters used to calculate each household's food security level, and (2) variability due to sampling error in aggregating across households in the CPS. In the estimates presented below, we take account of this second source of error but not the first. We will soon revise the estimation procedures to take account of both sources of error, using replication methods (see above) to reflect the CPS's complex design.

Weighting the Data

Not all households in a CPS sample have the same probabilities of selection. In deriving model parameter and insecurity prevalence estimates, we have weighted households to reflect their differential selection probabilities (and the effects of unit nonresponse adjustment and post-stratification), using the weights on the CPS files. Not using the weights does not clearly bias the estimates, since the underlying theory of the Rasch model does not require that the data set used to estimate parameters be representative of the population from which it was drawn. Indeed, item severity levels are explicitly defined to be independent of the estimation sample. Nevertheless, it seems prudent to determine whether our estimates are sensitive to whether weights are

[10] 18 positive scores plus zero.

used. Even if they are not, weights may be used to enhance the face validity of the work, given that most of the analysis of CPS data use the sampling weights.[11]

Standardizing the Rasch Scale

As noted earlier, a Rasch scale is uniquely determined only up to a linear transformation. That is, without loss of information, a specific Rasch scale can be rescaled so that the estimated parameters (or some subset of them) have any desired mean and standard deviation. Alternatively, the metric of a Rasch scale can be determined by anchoring any two parameters (such as the lowest and highest item severity levels) at any desired numerical values. The available software packages for estimating Rasch models use varying ways of normalizing the results they report.

In parts of the analysis—particularly those involving comparisons of item parameters across years—we have reported our results just as they come from the software that we are using to minimize the possibility (and the appearance) of inadvertently affecting comparisons through our choice of standardization. In other parts of our analysis, we have drawn on results based on a transformed food insecurity measurement metric ranging from zero to 10 to make them comparable to results reported from past work.

Establishing Cutpoints For Years Other Than 1995

The Rasch model estimates food insecurity levels for households on the basis of a numerical scale that is, in principle, continuous. (Though, with a finite number of questions, only a limited number of actual places on the continuous scale are observed.) One important objective of the government's food insecurity research has been to translate scores on this continuous scale into a small number of discrete food insecurity categories. To make this translation, it is necessary to establish cutpoints on the continuous scale that define the category boundaries. This section summarizes how this was done in the original 1995 analysis and then describes an approach that makes it possible to extend these methods to data for the later years.

Procedures Originally Used in Analyzing the 1995 Data

In the original analysis of the 1995 CPS data for categorizing households into discrete food security levels, the procedures used began by arraying the 18 food security questions in order of severity as estimated using the Rasch model. Then, based on the substance of the questions, the researchers, together with FNS, judgmentally assessed the seriousness of the food insecurity levels associated with modal sequences of answers. For instance, it was judged that, given the nature of the relevant questions, a household with children that answered the first 13 questions affirmatively should appropriately be placed in the most severe category of insecurity, while one that answered the first 12 questions affirmatively should be placed in the second most severe category. Thus, a cutpoint was established between the 12th and 13th question for the complete-data households with children. For this group of households (those that include children and have no missing data), all households answering 13 or more questions affirmatively were assigned to the most severe hunger category; conversely, those answering slightly fewer were assigned to the next less severe category.[12]

Table 3 summarizes the cutpoints that were established in this way. All households with complete data were assigned to one of the food insecurity categories shown in the table on the basis of how many affirmative answers they gave to the 18 questions, or 10 questions in the case of households without children.

[11]One other issue related to weighting should be noted. The weights on the 1995 file are incorrect, due to a Census Bureau processing error. We are attempting to obtain corrected weights, and we hope to use the corrected weights in later reports of our work.

[12]The approach draws on the characteristic of Rasch models, noted earlier, that scores for households with no missing data are uniquely determined by the number of affirmative answers; which of the questions have been answered affirmatively does not affect the score.

By itself, table 3 only applies to households with complete data. The method used to assign households with incomplete data to food insecurity categories was directly based on the numerical food security scores assigned to those households by the Rasch model. Implementing this procedure required establishing numerical cutoff values to define the borders of each food insecurity category. In doing this, the numerical cutoff between each adjacent pair of food security categories was set at approximately the level that separated cases with complete data in the two categories. (Remember, these cases with complete data were classified based on their numbers of affirmative answers, as described in the previous paragraphs.)

To illustrate this, we will continue to use the earlier example from the Hamilton et al. analysis. In that analysis, complete-data households with 13 affirmative answers, who were placed in the most severe hunger category on the basis of their number of affirmative answers, were assigned a food insecurity level of approximately 6.8 by the Rasch model, while those with 12 affirmative answers, who were placed in the next less severe category on the basis of their affirmative answers, were assigned a score of 6.4 by the Rasch model. Therefore, the numerical cutpoint between these categories used for households without complete data was a point between these two values, i.e., 6.4 and 6.8.

Procedures Used to Set Cutpoints for the 1996 and 1997 Data

In parts of our analysis, it has been necessary to establish cutpoints for the 1996 and 1997 data. This raises a number of complex issues, as discussed here.

Issues. One possibility for setting cutpoints for the 1996 and 1997 data is simply to take the numerical cutpoints for the 1995 data established in Hamilton et al. (1997) and to directly apply them to the households in the 1996 and 1997 data sets, including those with complete data and those without complete data. Several issues led us to reject this approach, however. One problem with doing this is that this approach logically requires that the Rasch model parameters be fully normalized so that they are on a comparable basis across the 3 years. While we are exploring ways of doing this, the most useful normalization to use in such work is not yet clear.

Perhaps more importantly, use across different years of fixed numerical cutpoints could potentially lead to measurement instability, caused by cross-year shifts of households between food insecurity categories as a result of very minor variations in the Rasch scoring. In particular, cutpoints in the previous analysis are very close to the numerical scores assigned by the model to large clusters of households. For instance, the numerical score assigned to all households with complete data and 13 affirmative answers is located only slightly above (in terms of severity) the numerical cutpoint established with the 1995 data as the boundary between the two most severe hunger categories. If the numerical cutpoints were kept the same across years and if all households were assigned to food security categories according to the cutpoints, then a slight change in model parameters between years could potentially cause the score assigned to this cluster of households to fall below the cutpoint into a different food insecurity category in the second year, resulting in considerable apparent instability in prevalence measures.[13]

How we have assigned the 1996 and 1997 cutpoints. To avoid the instability problem described above and preserve the basic logic under which categories were initially defined in the 1995 analysis, we have focused the cutpoints for observations with complete data on the numbers of affirmative answers, rather than on

[13]This instability issue pertains largely to households with complete data (10 or 18 answers, depending on the presence of children), since they form large clusters of observations in the data sets with identical scale scores. Households with incomplete data are more evenly distributed along the scale, because of the large number of permutations in which missing data can occur.

their numerical model scores.[14] In particular, all households with complete data have been assigned to food security categories on the basis of the decision rules summarized in table 3. This is essentially the exact procedure used for complete-data households in 1995, and it thus ensures comparability with that earlier work. Furthermore, as is reported below, our analysis of the 1996 and 1997 data suggests that the ordering of the severity of the items remains very similar across years, which lends further support to preserving the logic of the earlier analysis.

Households with incomplete data in the 1996 and 1997 data sets have been assigned to categories on the basis of their Rasch model numerical measures. New numerical cutpoints have been defined for each year. These new cutpoints have been defined in ways that are analogous to the approach used in 1995, except that they are based on the 1996 and 1997 model parameters. In particular, the cutpoints have been set such that they are the numerical values that separate different categories of households with complete data for the relevant years.[15]

Summarizing the above, the sequence of steps is essentially the following: households with complete data are being assigned to food insecurity categories, based on their numbers of affirmative answers to the CPS questions. Once those complete-data households have been assigned to categories, their scores are used to determine numerical scale levels, or cutpoints, that divide the various categories. Those numerical cutpoints are then used to assign households with incomplete data to categories.

[14]In the text, when we refer to "complete" data, we mean households with children that have 18 responses and households without children that have 10 responses. Technically, the households without children are treated as having missing data when estimating the Rash model. However, it is useful to think of them as complete for the discussion in the text, because they have no individual item nonresponse, and they, therefore, form clusters of households with the same values on the Rasch scale.

[15]The simpler alternative of keeping the 1995 cutpoints for use in assigning categories to households with incomplete data in the later years was rejected because of the danger that the 1995 values could become "out of line" over time with the Rasch scores for the households with complete data. In fact, as described below, over the 3 years currently under analysis, there is no appreciable difference in the results depending on whether the 1995 cutpoints or the updated cutpoints are used.

Correspondence With Earlier Findings

This paper draws on data for 3 years: 1995, 1996, and 1997. The data set we have used for 1995 is the same as that used in the earlier analysis, a fact that we have verified by fully reproducing selected key tables from Hamilton et al. (1997). However, the 1995 results reported in this paper differ slightly from those presented in Hamilton et al. The reasons for this include the following: (1) in most of the analysis we have imposed common screens to make the samples comparable across year, which had the effect of excluding some cases from the full 1995 data set; (2) in parts of our analysis, it has been convenient to report results using different Rasch model normalization conventions from that used in the earlier analysis; and (3) in parts of the analysis, we have used different conventions as to how cases with missing data are treated. While none of our 1995 results are in any way materially different from those reported in Hamilton et al., all or most of the exact numbers vary for the above reasons.

Software Used

Various software packages are available for estimating IRT models. These packages, while basically performing the same functions, often vary considerably along a variety of dimensions, including reporting formats, estimation algorithms used, statistical fit data reported, treatment of weighted data, and other features. Two of the available packages have been used in various parts of the analysis reported here, Bigsteps, which is maintained by Mesa Laboratories at the University of Chicago, and BILOG-MG, which is distributed by Scientific Software International. Various parts of the analysis reported below have been conducted with one or the other of these packages. Prior to our decision to use them interchangeably, we confirmed that both packages yield essentially identical estimates of basic model parameters, once differences in normalization metrics are taken into account.

Key Preliminary Findings

This section presents preliminary findings concerning the research questions highlighted in the introduction:

- Are estimated model parameters stable over time?
- How is the prevalence of food insecurity in the U.S. population changing?
- How robust are prevalence estimates to judgmental choices over alternative methods?

In addressing these questions and, in particular, the third question, we often present alternative model parameter or prevalence estimates, reflecting different methodological choices.

It should be emphasized that analysis of the data is still ongoing, and it is possible that some results presented may be revised, as additional analysis results become available.

Are Estimated Model Parameters Stable Over Time?

An important issue in examining the validity of the Rasch modeling approach is whether the model parameter estimates are stable over time. The underlying theory on which the Rasch model is based posits: if the wording of an item does not change, its estimated severity level should not change. Even if food insecurity became more prevalent over time, for example, a household at a given level of insecurity this year is assumed to answer each item the same as a household at that level of insecurity a year ago. Because of sampling variability and other factors, such as minor wording changes, we do not expect estimated model parameters to remain exactly the same over time, but a finding of major changes over time would call into question the validity of the model. Particularly problematic would be a finding of important changes in the ordering of the items by severity.

Table 4 shows estimated item parameters based on separate estimation, using data from each of the 3 years. (In estimating these models, we have used the tighter common screen, and households with missing data are retained in the analysis.) To avoid the possibility of influencing the results by imposing a normalization on the data, we have chosen in this table to present the model results directly as they were printed by the software program used.

These model estimates show considerable stability of model parameters across years. Particularly important, the order of the 18 items, in terms of their estimated severity, remains completely constant over the 3 years.[16]

As shown in table 4, the item severity estimates vary somewhat over time, but the degree of variation is quite small, relative to the overall range of the scale that extends about 3.5 units. For instance, the severity of the most severe (and least precisely estimated) item, item 50, fluctuates only slightly. From a value of 3.01 in 1995, it drops to 2.89 in 1996 and rises to 3.07 in 1997. Fluctuations are greater for some other items. The largest difference over time is for item 47, which rises from 1.85 to 2.08 between 1995 and 1997, an increase of 0.23.

In assessing these results, it is important to note that, as indicated above, we have not normalized the scales in any way to keep them comparable over the 3 years. Some differences across years in parameters may be an artifact of different implicit normalizations. For instance, all but 2 of the 18 scores in the 1997 data are higher than those in 1995. If we had chosen to anchor one of the 1997 items to its 1995 value, it is likely that the degree of apparent variation in the scores would have been lessened.[17]

[16]In other variants of these analyses, such as those with different screens, there was some tendency for one pair of adjacent items, items 28 and 40, to invert their order in different years. However, this inversion does not significantly affect other components of the analysis, and after examining many variants of the estimated results, our conclusion is that the item ordering is highly stable across years.

[17]We anticipate that in later work from the project, some normalization will be used in reporting cross-year scale comparisons. We are still examining the issue for the most appropriate normalization for this purpose.

Even though the changes in parameters are relatively small, many of them are statistically significant, in part because of the very large sample sizes in the CPS. A typical estimated standard error in the middle or bottom part of the table is about 0.03. With this amount of sampling variation, differences across years in item severity levels of 0.08 or more are generally statistically significant. However, two points should be noted: (1) while we are still examining implications of these cross-year differences in item severity, our preliminary assessment is that differences of the magnitude shown in table 4 will not have any material effect on prevalence estimates; and (2) as discussed above, once a common normalization is imposed on the results in table 4, it is likely that the sizes of the cross-year differences and their statistical significance will be reduced.

Table 5 examines the robustness of the item severity results to differences in assumptions concerning screening and the treatment of missing data. For illustration, the table focuses on 1997 data, but similar results are found when data for other years are examined. The table shows that using the loose screen rather than the tight screen results in changes in item severity roughly comparable in magnitude to the cross-year changes observed in the previous table. The choice as to whether to use observations with missing data has virtually no impact. (This latter result is not surprising, given the very low incidence of item nonresponse in the data set.)

How is the Prevalence of Food Insecurity in the U.S. Population Changing?

Final decisions about how to develop estimates of changes in food security levels over time are still under consideration. Key issues include:

- Should model parameters be re-estimated each year?

- Should some sort of intertemporal averaging be used to derive parameter estimates?

To provide a preliminary look at changes in the prevalence of food insecurity over time, we have developed time series estimates based on the 1995 parameter estimates. Specifically, we have applied the 1995 item severity parameters and the 1995 cutpoints to each of the 3 years of data. In addition, we have made preliminary prevalence estimates based on model parameters derived from the 1996 and 1997 data. The results are reported here.

Prevalence estimates based on 1995 model parameters show a noticeable increase in food security over time (table 6). In particular, after increasing by 0.1 percentage point between 1995 and 1996, the percentage of the population classified as food secure increases by a substantial and statistically significant 1.6 percentage points between 1996 and 1997.[18] As indicated in the lower rows of the table, the overall increase of 1.7 percentage points in the percentage of households classified as food secure is reflected in decreases in each of the three levels of food insecurity. Between 1995 and 1997, households categorized as food secure without hunger decrease by 0.8 percentage point, while the two groups classified as experiencing hunger drop by 0.7 and 0.2 percentage point.

Table 7 examines the sensitivity of these findings to alternative methodologies, using the 1997 data for illustration. Applying the loosest common screen rather than the tighter common screen decreases by 0.6 percentage point the percentage of the population classified as food secure. The direction of the effect is expected, because applying the looser screen allows some households to be classified based on their responses to the full battery of food security questions rather than simply being classified as food secure based on the screening questions. Not surprisingly, almost all of the effect of using the looser screen is to transfer cases from the food secure category on to the food insecure without hunger category. The percentages in the two most severe food insecurity categories remain essentially the same.

[18]Work estimating the standard errors of the prevalence estimates in the table is underway but has not yet been completed. However, preliminary calculations suggest that the substantial 1996-97 change discussed in the text is almost certain to prove statistically significant.

The exclusion of households with missing data results in 0.5 percent of households being transferred out of the other categories and into a category of status not determined. About half of the households whose status is changed come from the food secure category. Most of the others are transferred from food insecure without hunger and food insecure with hunger.

Whereas tables 6 and 7 focus on the effects of alternative methodological choices on the estimated prevalence levels, table 8 focuses on the effects of alternative methodologies on the estimated changes in prevalence levels over time. The first column reproduces the 1995-97 change estimates reported in table 6, while the second and third columns show estimated changes under the alternative methodologies that we are examining. In general, only slight variation is observed, and neither of the alternative methodologies being examined appears to substantially affect estimates of changes in prevalence.

Prevalence Estimates Based on 1996 and 1997 Parameters

All of the prevalence analysis up to this point has been based on applying the 1995 model parameters and classification rules to all 3 years of data. As a preliminary step in developing recommendations about how to update the analysis methods over time, we have also experimented with using 1996 and 1997 model parameters separately to estimate food insecurity prevalence levels for those 2 years.

The first two columns of table 9 display 1996 prevalence estimates based on alternative sets of Rasch parameters. The first column reproduces the 1996 prevalence based on 1995 parameters as reported earlier in table 6. The second column shows results based on the methods previously described that draw, in part, on Rasch model parameters estimated with 1996 data. The changes from using the 1996 parameters are minimal. None of the prevalence percentages change by more than 0.1 percentage point.

Comparable data are also presented in the table for the 1997 data. Again, there are no substantial changes in the estimated prevalence rates.

In assessing these results, it should be noted that for more than 97 percent of households with complete data, there is essentially no difference between the analysis in this section and the analysis based fully on 1995 parameters. This is because under the Hamilton et al. analysis approach, as replicated in earlier subsections, households with complete data are essentially being classified based on their numbers of affirmative answers, and this approach is not being changed in the results reported above. Thus, it is essentially only the treatment of households with missing data that can be affected by the new methods being examined in this section, which allows the numerical Rasch model scores to be separately estimated for 1996 and 1997. Further, even for these households, their classification into food insecurity categories is unlikely to be affected unless they are quite near a margin between categories.

Other Empirical Results

In addition to the central findings reported above, several other methodological questions have been examined. Selected issues are discussed in this section.

How Modal are Household Response Patterns?

The Rasch model implies that most households will exhibit item response patterns that are reasonably modal, in the sense that if a household answers “yes” to any of the items, it will tend to answer “yes” to the least severe items first, and then answer “no” to the more severe items. A household that exhibits this pattern exactly—a string of all “yes” answers followed by a string of all “no” answers—is said to be a “modal” household. There is nothing in the Rasch theory that predicts that all households will be modal; indeed, the model cannot be estimated if all households are exactly modal. Still, it is of interest in understanding the data to examine the degree of modality that is present. Large numbers of strongly non-modal response patterns could call into question the validity of the model.

One approach to examining the degree to which households exhibit modal answers is to calculate for each household the minimum number of answers that would have to be different in order to make the household responses be modal. Of course, if the household's answers are already perfectly modal, then the number of answers that would have to be changed is zero. However, consider, as an example, a household with the following response pattern: three "yes" answers, then a "no," then a "yes," then all "no's." For such a household, only one item (the first "no" or the last "yes") would have had to be changed to make the response pattern modal. Similarly, to take a second example, suppose a household has two "no" answers, four "yes" answers, then all "no" answers. It would require at least two changes (the first two "no" answers) to make the household modal.

Table 10 tabulates the minimum number of answer changes required to make the households in the 1995 sample modal. It shows that 37 percent of the households in the 1995 data are perfectly modal. For another 36 percent, there is only one discrepancy between their scores and a modal pattern. Sixteen percent have two such discrepancies, and 10 percent have three or more. Overall, this suggests a pattern of substantial adherence to modal response patterns.[19]

Households without children appear to exhibit more modality than those with children. However, this may be due to the fact that there are fewer questions applicable to the group without children (10 rather than 18 for households with children) and, hence, fewer opportunities for non-modality.

Table 11 presents a more detailed look at these issues, focusing on households with children. The central section of the table shows, for each possible number of "yes" answers, the frequency distribution of the highest item (in terms of severity) to which the non-modal households with children answered "yes." For instance, the fifth row shows data for the 335 non-modal households that gave five "yes" answers. For 119 of them, the highest "yes" answer was on item 6, while for another 75, the highest "yes" answer was item 7, and so on.

The shading in the table reflects the fact that certain cell entries are logically impossible—if, for instance, there are five "yes" answers, the highest non-modal item with a "yes" answer cannot come before the sixth item. To the extent that the non-modal households are almost modal, we would expect households to be clustered just to the right of the shaded area. For instance, using the previous example, a household with five "yes" answers that has the sixth item as its highest "yes" answer has only one non-modal answer in its overall string of answers. For most rows in the table, non-modal households do cluster near the shaded diagonal, suggesting that the non-modal response patterns are not severely non-modal. About 50 percent of households are in the first two off-diagonal cells, and an additional 20 percent are in the third cell.

Bounds on the Effects of Non-Modality on Prevalence Estimates

A useful way to understand the implications of non-modal response patterns is to assess their effects on prevalence estimates. Accordingly, in this section, we calculate for each household the minimum and maximum food insecurity levels that can be obtained by making the household's response pattern modal. To obtain the minimum insecurity level, we classify a household based on the items before its first "no" answer. This effectively converts all higher "yes" answers to "no" answers, giving a modal pattern. To obtain the maximum insecurity level, we classify a household based on the items up to and including its last "yes" answer. This effectively converts

[19]Of course, for a household with just one non-modal answer, that answer could be severely non-modal (a "no" at the beginning of a long string of "yes" answers or a "yes" many items after the previous "yes"). A little later, we will examine the severity of non-modality.

all lower "no" answers to "yes" answers, again giving a model pattern.

Consider, as an example, a household with children that answers "yes" to the first 2 items, "no" to the third and fourth items, "yes" to the next 4 items, and "no" to the last 10 items (for six "yes" answers in all). For this household, the minimum food insecurity level is based on the modal pattern of "yes" to the first two items and "no" to the last 16 items. The maximum food insecurity level is based on the modal pattern of "yes" to the first 8 items and "no" to the last 10 items.[20]

When every household is classified at its minimum food security level, the overall prevalence of insecurity (in the highest category and across the three categories) is at a minimum. Likewise, when every household is classified at its maximum food insecurity level, the overall prevalence of insecurity is at a maximum.

The results of this analysis are reported in table 12. In each of the 3 years, going to the "minimum insecurity" scenario tends to increase the estimated proportion of food secure households by about 1 percentage point, compared with the base estimates, and there is a decrease in the proportion classified as experiencing hunger of between 1 and 2 percentage points. Going to the scenario with "maximum insecurity" raises the proportion with hunger by between 2 and 5 percentage points, depending on the year.

Interestingly, the category that involves the minimum food insecurity estimate causes the proportion of households in the middle category—food insecure without hunger—to be higher in each of the 3 years. This is because more households move into this category from the most severe category than leave it to go into the food secure category.

[20]In contrast, the Rasch model would assign this household the same insecurity score as any other household with 6 "yes" answers (and no missing responses), effectively treating it as though it had the modal response pattern of "yes" to the first 6 items and "no" to the last 12.

Conclusions

This paper has summarized selected preliminary results from work that is still very much in progress. It is possible that some results we have presented could change significantly after further analysis and review, and additional lines of analysis remain to be carried out. However, it may be useful to conclude by summarizing key findings to date. They include:

- The Rasch model parameters do not change greatly, depending on which data set is used to estimate the model; importantly, the item ordering remains approximately constant.
- It has been possible to identify two data screens that place the data from the 3 years on a comparable basis; results do not appear highly dependent on which of these screens is used, or indeed on whether any screen is used.
- The prevalence of food insecurity appears to have declined between 1995 and 1997.
- It is useful in applying the model to other years to rely heavily on the relationships between numbers of affirmative answers and food insecurity categorization developed in the original analysis of the 1995 data. By doing this, it has been possible to develop an approach that appears to obtain reasonable prevalence results, even using models estimated from different years' data sets.

In addition to refining the analysis done to date, we anticipate that we will:

- Develop appropriate ways to estimate the substantive and statistical significance of changes in prevalence estimates over time,

- Finalize decisions concerning what screen to use and how to treat missing data,

- Finalize decisions as to how to establish model parameters and cutpoints for longitudinal analysis,

- Develop appropriate procedures for normalizing Rasch scales estimated with data from different years,

- Apply the analysis to the 30-day data, and

- Conduct analysis of subgroups of the overall household population.

References

Hamilton, W.L., J.T. Cook, W.W. Thompson, L.F. Buron, E.A. Frongillo, Jr., C.M. Olson, and C.A. Wehler. 1997a. *Household Food Security in the United States in 1995: Summary Report of the Food Security Measurement Project.* U.S. Department of Agriculture, Food and Consumer Service (currently Food and Nutrition Service), Office of Analysis and Evaluation, Alexandria, VA.

Hambleton, R.K., and R.W. Jones. Fall, 1993. "An NCME Instructional Module on Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development," *Educational Measurement: Issues and Practice*. Vol. 12, No. 3. pp. 38-47.

Radimer, K.L., C.M. Olson, J.C. Greene, C.C. Campbell, J.P. Habicht. 1992. "Understanding Hunger and Developing Indicators to Assess It in Women and Children," *Journal of Nutrition Education*. Vol. 24. pp. 36S-45S.

Wehler, C.A., R.I. Scott, J.J. Anderson, L. Summer, and L. Parker. 1995. *Community Childhood Hunger Identification Project: A Survey of Childhood Hunger in the United States.* Food Research and Action Center, Washington, DC.

Wright, B.D., and G.N. Masters. 1982. *Rating Scale Analysis.* Chicago: MESA Press.

Table 1—CPS sample sizes, unweighted number of households

Item	April 1995	September 1996	April 1997
		Number	
Full CPS	53,665	47,795	47,306
Households in supplement	44,730	41,811	41,146
Households tracked into food security module[1]	18,453	10,957	11,175
Answered all key questions asked	18,179	10,685	10,937
Answered at least half of key questions asked, but not all	195	203	171
Answered fewer than half of key questions asked	79	69	67

[1]There are 18 key questions for households with children and 10 for those without children.

Table 2—Sample sizes under alternative screens

Item	Low-income households	High-income households	Total
		Number	
1995:			
Maximum sample available	15,662	2,791	18,453
Loosest screen[1]	7,891	2,791	10,682
Tighter screen[2]	5,049	2,791	7,840
1996:			
Maximum sample available	7,259	3,698	10,957
Loosest screen[1]	7,259	2,674	9,933
Tighter screen[2]	4,760	2,674	7,434
1997:			
Maximum sample available	6,293	4,882	11,175
Loosest screen[1]	6,293	2,640	8,933
Tighter screen[2]	4,084	2,640	6,724

[1]See the Loosest Common Screen box.
[2]See the Tighter Common Screen box.

Table 3—Food insecurity status by number of affirmative answers, households with complete data

Food insecurity status	Affirmative answers: Households with children	Households without children
	Number	
Food secure	0-2	0-2
Food insecure without hunger	3-7	3-5
Food insecure with moderate hunger	8-12	6-8
Food insecure with severe hunger	13-18	9-10

Table 4—Item parameter estimates based on cases that pass the tighter common screen, includes complete cases and all cases with item non-response, 1995-97

		1995		1996		1997	
Item	Description	Parameter	Standard error	Parameter	Standard error	Parameter	Standard error
NHES50	Child not eat for whole day	3.01	0.11	2.89	0.07	3.07	0.11
NHES44	Child skipped meal, 3+ mos.	2.51	.06	2.43	.09	2.47	.09
NHES43	Child skipped meal	2.28	.05	2.22	.07	2.22	.08
NHES29	Adult not eat whole day, 3+ mos.	2.02	.05	1.98	.03	2.13	.04
NHES47	Child hungry	1.85	.04	1.95	.04	2.08	.07
NHES28	Adult not eat whole day, 3+ mos.	1.78	.04	1.77	.06	1.91	.05
NHES40	Cut size of child's meals	1.76	.04	1.72	.02	1.89	.03
NHES38	Adult lost weight	1.68	.03	1.55	.04	1.74	.03
NHES35	Adult hungry but didn't eat	1.19	.02	1.14	.04	1.33	.03
NHES57	Child not eating enough	1.10	.03	1.10	.03	1.22	.05
NHES25	Adult cut size or skipped meals, 3+ mos.	.82	.03	.80	.02	.96	.03
NHES56	Adult eat less than felt they should	.53	.03	.52	.04	.73	.04
NHES32	Couldn't feed child balanced meals	.49	.03	.45	.02	.62	.03
NHES24	Adult cut size or skipped meals, 3+ mos.	.44	.03	.40	.02	.55	.02
NHES58	Adult fed child few low-cost foods	.03	.03	.04	.03	.20	.03
NHES55	Adult not eat balanced meals	-.12	.02	-.13	.02	.05	.03
NHES54	Food bought didn't last	-.25	.03	-.24	.02	-.14	.02
NHES53	Worried food would run out	-.53	.03	-.57	.02	-.48	.02

Note: Standard errors are computed using a balanced repeated replication jackknife procedure to account for the complex sample design used in the Current Population Survey. These estimates are based on weighted data.

Table 5—Sensitivity of item severity estimates to different assumptions, 1997 data

		Estimated item severity levels		
Item	Description	Basic analysis[1]	With looser screen	Excluding all cases with item nonresponse
NHES50	Child not eat for whole day	3.07	3.13	3.08
NHES44	Child skipped meal, 3+ mos.	2.47	2.56	2.48
NHES43	Child skipped meal	2.22	2.32	2.23
NHES29	Adult not eat whole day, 3+ mos.	2.13	2.23	2.14
NHES47	Child hungry	2.08	2.20	2.09
NHES28	Adult not eat whole day, 3+ mos.	1.91	2.04	1.92
NHES40	Cut size of child's meals	1.89	2.01	1.90
NHES38	Adult lost weight	1.74	1.87	1.74
NHES35	Adult hungry but didn't eat	1.33	1.51	1.34
NHES57	Child not eating enough	1.22	1.36	1.23
NHES25	Adult cut size or skipped meals, 3+ mos.	.96	1.17	.97
NHES56	Adult eat less than felt they should	.73	.91	.74
NHES32	Couldn't feed child balanced meals	.62	.84	.62
NHES24	Adult cut size or skipped meals, 3+ mos.	.55	.79	.56
NHES58	Adult fed child few low-cost foods	.20	.43	.21
NHES55	Adult not eat balanced meals	.05	.29	.05
NHES54	Food bought didn't last	-.14	.15	-.14
NHES53	Worried food would run out	-.48	-.22	-.48

[1]Reproduced from table 4.

Table 6—Food insecurity prevalence estimates by year with severity levels based on 1995 data

Food security status	1995[1]	1996	1997	Change in estimates, 1995-97
	Percent			*Percentage points*
Food secure	89.4	89.5	91.1	1.7
Food insecure without hunger	6.4	6.2	5.6	-.8
Food insecure with hunger	3.3	3.3	2.5	-.8
Food insecure with severe hunger	.8	.9	.6	-.2
Food security status not determined	.2	.1	.2	0

Note: Estimated for each year, using model parameters based on 1995 data. Based on the tight screen and inclusion of all cases with missing data.

[1]Estimates differ from those published in Hamilton et al. (1997) because the screens used to track households into the detailed food security analysis have been adjusted to make them consistent across the 3 years.

Table 7—Effects of alternative methodologies on estimated prevalence levels, 1995 data

Food security status	Main analysis[1]	Excluding cases with missing data	Using looser screening criteria
		Percent	
Food secure	89.4	89.2	88.8
Food insecure without hunger	6.4	6.3	7.1
Food insecure with hunger	3.3	3.2	3.3
Food insecure with severe hunger	.8	.8	.8
Food security status not determined	.2	.5	—

[1]From table 6.

Table 8—Effects of alternative methodologies on estimated changes in prevalence levels, 1995-97

Food security status	Main analysis[1]	Excluding cases with missing data	Using looser screening criteria
	Percentage points		
Food secure	1.7	1.7	1.8
Food insecure without hunger	-.8	-.9	-.9
Food insecure with hunger	-.8	-.8	-.8
Food insecure with severe hunger	-.2	-.2	-.2
Food security status not determined	—	.2	—

[1]From table 6.

Table 9—Prevalence estimates based on 1996 and 1997 model parameters

Food security status	Prevalence estimates			
	1996 data		1997 data	
	1995 parameters[1]	1996 parameters	1995 parameters[1]	1997 parameters
	Percent			
Food secure	89.5	89.6	91.1	91.2
Food insecure without hunger	6.2	6.2	5.6	5.7
Food insecure with hunger	3.3	3.3	2.5	2.5
Food insecure with severe hunger	.9	.9	.6	.6
Food security status not determined	.1	—	.2	—

Note: Based on tight screen and inclusion of all cases with missing data.

[1]From table 6.

Table 10—Percentage of households by the minimum number of non-modal responses to the food security items, 1995 data

Non-modal responses[1]	All households	Households without children	Households with children
		Percent	
0	37	49	28
1	36	37	34
2	16	10	21
3	7	3	10
4	2	1	4
5	1	0	2
6+	0	0	1
Total	100	100	100

[1]Minimum number of responses that would have to be changed for the household to be modal.

Table 11—Analysis of modality by number of "yes" responses to food security items, unweighted households with children, 1995

	All		Modal		Non-modal - highest "yes" response item																	Non-modal	
Number of "yes" responses	HH[1]	Percent	HH	Percent	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	HH	Percent
1	970	22.6	583	48.6	107	82	131	26	4	24	0	3	3	1	3	2	0	0	0	0	1	387	12.5
2	661	15.4	273	22.8		84	177	22	32	34	19	8	6	1	0	0	3	1	1	0	0	388	12.5
3	550	12.8	113	9.4			171	37	86	78	24	18	8	2	3	2	7	0	1	0	0	437	14.1
4	386	9.0	70	5.8				29	101	84	42	29	11	8	6	3	2	0	1	0	0	316	10.2
5	343	8.0	8	.7					119	75	60	34	22	7	4	3	3	4	2	2	0	335	10.8
6	358	8.3	14	1.2						63	84	127	37	7	10	4	7	4	1	0	0	344	11.1
7	255	5.9	17	1.4							65	57	43	21	13	15	11	8	2	2	1	238	7.7
8	188	4.4	38	3.2								31	33	28	14	13	12	12	4	1	2	150	4.8
9	176	4.1	35	2.9									31	21	28	22	18	15	2	2	2	141	4.6
10	132	3.1	20	1.7										14	24	12	28	22	4	6	2	112	3.6
11	86	2.0	6	.5											14	3	23	28	7	5	0	80	2.6
12	59	1.4	3	.3												3	19	16	2	14	2	56	1.8
13	59	1.4	0	0													13	24	9	8	5	59	1.9
14	28	.7	0	0														14	2	11	1	28	.9
15	15	.3	4	.3															0	8	3	11	.4
16	12	.3	0	0																7	5	12	.4
17	13	.3	9	.8																	4	4	.1
18	6	.1	6	.5																		0	0
Total	4,297	100.0	1,199	100.0	107	166	479	114	342	358	294	307	194	110	119	82	146	148	38	66	28	3,098	100.0
Percent					3.5	5.4	15.5	3.7	11.0	11.6	9.5	9.9	6.3	3.6	3.8	2.6	4.7	4.8	1.2	2.1	.9	100.0	

[1]HH is the abbreviation for households.

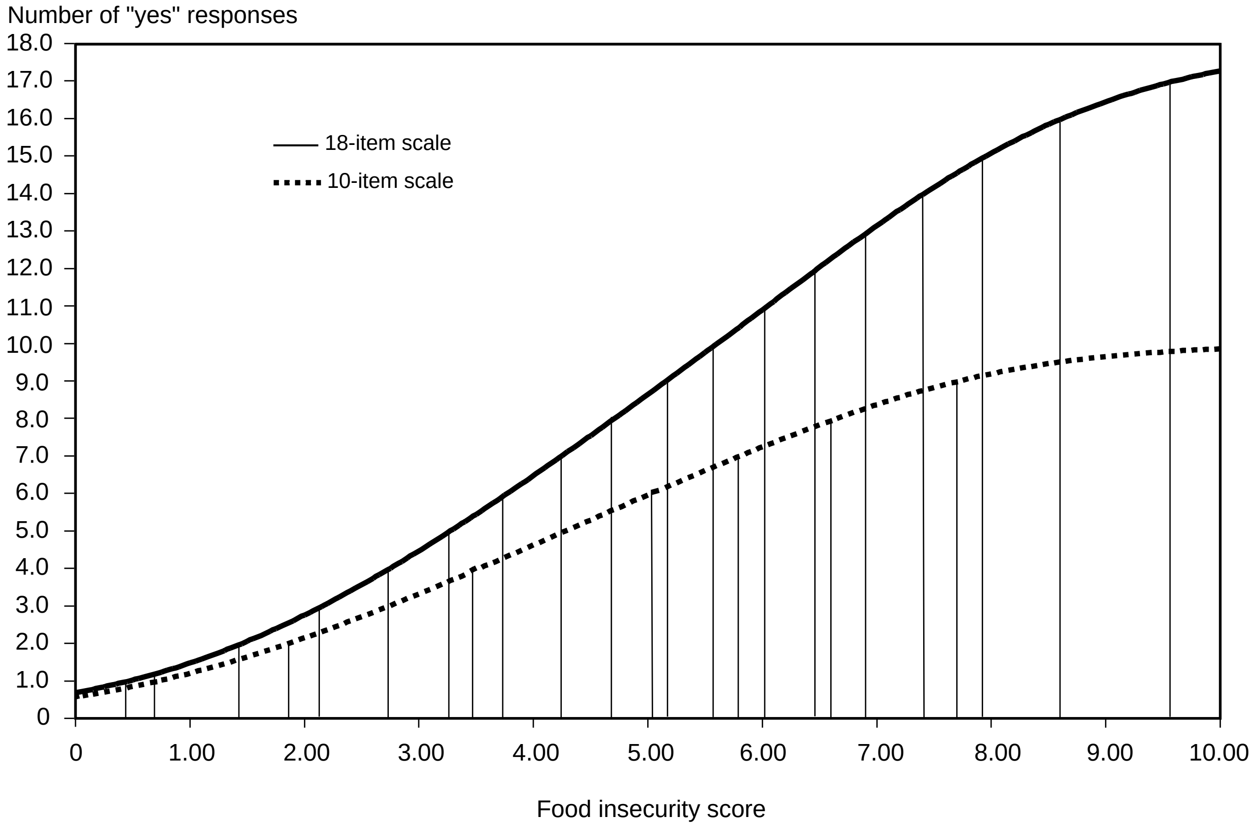
Table 12—Minimum and maximum food insecurity prevalence estimates, percentage of households, 1995-97

Food security status	Base estimate	Estimate involving minimum estimate of food insecurity	Estimate involving maximum estimate of food insecurity
		Percent	
1995:			
Food secure	89.4	90.33	87.16
Food insecure, hunger not evident	6.4	6.84	6.06
Food insecure, hunger evident	4.1	2.68	6.78
Food security status not determined	.2	.15	
1996:			
Food secure	89.5	90.42	87.13
Food insecure, hunger not evident	6.2	6.70	6.08
Food insecure, hunger evident	4.2	2.72	6.79
Food security status not determined	.1	.17	
1997:			
Food secure	91.1	91.95	88.14
Food insecure, hunger not evident	5.6	5.83	3.97
Food insecure, hunger evident	3.1	2.10	7.89
Food security status not determined	.2	.12	

Notes: To compute the maximum estimate of food insecurity, households were classified based on the most severe item with “yes” responses. To compute the minimum estimate of food insecurity, households were classified based on the most severe “yes” item preceding the least severe “no” response. Percentages may not add to 100 due to rounding.

Figure 1

18- and 10-item test characteristic curves, based on 1995 item calibrations as reported in Hamilton et al. (1997)



Note: Positions of vertical lines represent item scores.

Appendix
Key Aspects of the Rasch Model

As noted in the text, household answers to 18 questions from the Current Population Survey (CPS) supplements have been used to develop food insecurity scales. This appendix summarizes the scaling methods used.

As with most surveys, not all the CPS questions were asked of all households. Questions that were not applicable to a household were skipped during the interview. The most important example in the food security supplement is that questions about children were not asked of households with no children. While households with children answered 18 items, childless households answered only 10 items, skipping the 8 items pertaining to children. This creates a problem: how do we measure the food insecurity status of all households, both those with and those without children, on a common scale?

Item response theory (IRT) provides a solution to this problem. (See, for instance, Hambleton (1993) for a discussion of IRT procedures.) One way to understand how IRT deals with this problem is to consider a more traditional application of IRT models: student testing. The problem is found in testing because students are given different versions of a test to deter cheating, but all students need to be graded on a common scale. This issue arises with the Scholastic Aptitude Test (SAT) and other standardized tests, as well as in educational research studies, such as the National Assessment of Educational Progress.

The problem is resolved by first determining how difficult each test item is relative to every other item. In general, a more difficult item is answered correctly by fewer students than is a less difficult item. Difficulty can be quantified and estimated using maximum likelihood statistical methods. Difficulty levels are the item scores—or item parameters—in the IRT model.[1] The difficulty of the overall test can be expressed as a function of the difficulty of all items that comprise the test. This function is the test characteristic function, and a graph of the function is the test characteristic curve. A grade can then be assigned to each student based on the number of items answered correctly and the characteristic function for the version of the test taken. It is the use of the test characteristic function that adjusts each student's raw score (the number of correct responses) for the difficulty of the items in that student's version of the test. These grades are the respondent scores (or respondent parameters) in the IRT model. Taken together, these two pieces of information—the number of correct responses and the characteristic function of each student's test—allow all students to be graded on the same scale, even though they are given different tests. For example, a score of 8 out of 11 items on an easy test might turn out to be equivalent to a score of 6 out of 13 items on a more difficult test. The IRT model provides the theoretical foundation and the mathematical relationships needed to quantify the measures of item difficulty and student performance. The IRT model also provides the statistical theory needed to estimate the item parameters and student scores.[2]

In our application to food insecurity measurement, two different versions of a test are administered in the CPS. The first version, administered to households with children, has 18 items. The second version, administered to childless households, has only 10 of the 18 items. Just as different items on an educational test tap different levels of student ability—that is, some items are more difficult than others—each of the 18 CPS questionnaire items taps a different level of household food insecurity. To measure the food insecurity status of all households on the same scale and use all available information from the survey, we score households with children using

[1]There are a variety of IRT models in use in testing applications. The differences among them are primarily related to the number of parameters—typically, one to three—associated with each item. The work done by Abt with the 1995 CPS food security data and the work presented in this paper for all 3 years are based on a one-parameter IRT model, also known as a Rasch model.

[2]Neither item parameters nor student scores are unique; they are determined only up to a linear transformation.

the 18 items that they are asked, and score childless households using the 10 items that they are asked.[3]

Figure 1 illustrates how households were scored for the 1995 food security prevalence estimates (Hamilton et al., 1997).[4] The figure shows the test characteristic curves for the 18-item test (the dotted curve) and for the 10 item test (the dotted curve). The figure is based on the item parameters estimated from the April 1995 data. For measuring food insecurity, responses to each of the 18 items have been coded as "yes" or "no." A "yes" provides evidence of food insecurity. The process used to assign each household a food insecurity score is as follows:

- Count the number of "yes" responses to the 18 items—or to the 10 items, in the case of childless households.

- Find that number on the vertical axis in figure 1.

- For households with children, read across to the heavy black curve. For households without children, read across to the dotted curve.

- Read down to the horizontal axis. The value on the horizontal axis is the score assigned to the household.

For example, a household with children that answers "yes" to seven items would be assigned a score of about 4.25. In contrast, a childless household that answers "yes" to seven items would be assigned a score of about 5.75, substantially higher than the score of 4.25 assigned to the household with children. As it turns out, a childless household that answers "yes" to five items would be assigned roughly the same score as a household with children that answers "yes" to seven items, indicating that according to this one-parameter IRT model and the 18 CPS items, these two households are experiencing about the same levels of food insecurity.

In the current application, as we discussed earlier, there are actually more than two versions of the food security test, because of the many possible different patterns that can be observed in the missing data. The different scoring of households with and without children is an example of how the more general problem of missing data is handled in much of the analysis reported here. A test characteristic function can be defined for each household based on the set of items to which the household responded. For example, a household with children that had missing data for items 5 and 17 would be scored using a different test characteristic function than a household with children that had missing data for items 9 and 13. By using the appropriate test characteristic function for each household, all households are placed along a common food insecurity scale.

[3]Another option would be to use only the 10 items common to all households. This has the appeal of simplicity. However, doing this would mean disregarding a substantial amount of information about households with children.

[4]Figures follow tables at the end of this paper. This figure does not directly appear in the Hamilton et al. report but is based on parameters reported there.