United States Department of Agriculture

Economic Research Service

Technical Bulletin Number 1936

December 2012

USDA

Assessing Potential Technical Enhancements to the U.S. Household Food Security Measures

Mark Nord

www.ers.usda.gov

Visit Our Website To Learn More!

http://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us.aspx

Recommended citation format for this publication:

Nord, Mark. *Assessing Potential Technical Enhancements to the U.S. Household Food Security Measures.* TB-1936, U.S. Department of Agriculture, Economic Research Service, December 2012.

United States Department of Agriculture

Technical Bulletin Number 1936

December 2012

A Report from the Economic Research Service

www.ers.usda.gov

Assessing Potential Technical Enhancements to the U.S. Household Food Security Measures

Mark Nord, marknord@ers.usda.gov

Abstract

The statistical measures used by the U.S. Department of Agriculture since 1995 to monitor the food security of the Nation's households—the extent to which they can consistently acquire adequate food for active healthy living—are based on a single-parameter logistic latent-trait measurement model (the Rasch model). A panel convened, at USDA's request, by the Committee on National Statistics (CNSTAT) of the National Academies in 2003-06 recommended that USDA explore five potential technical enhancements to that model. USDA has adopted one CNSTAT panel recommendation, which corrects the methods used to model the frequency-of-occurrence followup questions in the food security scale. This study examines the implications of that change and assesses the other four potential enhancements and the extent to which they would affect USDA's published food security statistics. The study findings suggest that introducing the more complex statistical models would improve measurement of food security little, if at all, while making results and methods more difficult to explain to policy officials and the public.

Keywords: food security, food security measurements, food insecurity, Rasch model, item response theory

About the Author

Mark Nord is a sociologist in the Food Assistance Branch, Food Economics Division, Economic Research Service, U.S. Department of Agriculture.

Acknowledgments

Thanks to Helen Jensen and Amy Froelich, Iowa State University; Matthew Johnson, Columbia University Teachers College; David Holben, Ohio University; Megan Deitchler, FHI 360; Alisha Coleman-Jensen, USDA, Economic Research Service; and to the editor, Priscilla Smith, and designer, Wynnice Pointer-Napper, USDA, Economic Research Service.

Contents

Summary

What Is the Issue?

The U.S. Department of Agriculture's Economic Research Service (ERS) monitors the food security of the Nation's households—the extent to which they can consistently acquire adequate food for active healthy living—using data from an annual, nationally representative survey. Responses to multiple indicators of food insecurity by each surveyed household are combined to determine the food security status of the household. Statistical methods based on a single-parameter logistic latent-trait measurement model are used to assess the food security questions and scales based on them.

In 2003-06, at USDA's request, the Committee on National Statistics (CNSTAT) of the National Academies convened an expert panel to assess the methods USDA uses to measure and report household food security. The panel recommended that USDA continue to monitor food insecurity, affirmed the general statistical approach, and recommended that USDA consider several potential technical enhancements to the statistical methods. This report describes findings from ERS' assessments of five of those potential enhancements.

What Were the Study Findings?

- **It may not be appropriate to incorporate all available frequency-of-occurrence information into the main measure using polytomous item-response theory (IRT) statistical models.** The current standard measure represents the greatest severity of food insecurity experienced at any time during the year. Frequent or persistent food insecurity appears to represent a somewhat distinct dimension, and it may not be appropriate to represent the two dimensions in a single measure. An alternative may be to represent frequent or persistent food insecurity based on a separate scale.

- **Frequency-of-occurrence followup questions that are included in the measure should be modeled along with their base items as ordered polytomous items rather than as two independent questions.** ERS has already adopted this methodology as recommended by the CNSTAT panel.

- **Allowing item-discrimination parameters to differ from item to item would improve measurement precision only slightly and would make prevalence statistics less understandable to a lay audience than those based on the single-parameter model.**

- **The extent of differential item function (DIF) between households with and without children is not great enough to substantially distort comparisons of prevalence.**

- **Assigning the food security status of households probabilistically to reflect the measurement error inherent in the latent-trait measurement model would reduce the error in prevalence estimates and eliminate the bias in the current methodology between households with and without children, but would not change other patterns of prevalence across subpopulations or over time.** Although the bias between households with and without children is an important issue, that bias could,

alternatively, be obviated by cross-classifying households with children based on separate measures of food insecurity for adults and children. Such an approach would be more readily understood by policy officials and the public and less disruptive to the overall measurement system.

- **The findings hold when three methods are assessed in combination.** Patterns of food security prevalence over time and across subpopulations based on the most complex model—combining three of the possible enhancements—differ little from those based on the current methods.

The findings suggest that little would be gained by measuring food security with any of the more complex measures, provided an alternative methodology can be implemented to remove the current bias in comparing the prevalence of food insecurity between households with and without children.

How Was the Study Conducted?

The various measurement models were explored using data from the Current Population Survey Food Security Supplements (CPS-FSS). ERS sponsors the annual collection of the CPS-FSS, which is conducted by the U.S. Census Bureau and includes 45,000 to 50,000 households each year. The survey provides the data for USDA's annual report on food security in U.S. households. Data for various years and multiyear periods from 1995 to 2010 were used for the analyses.

Introduction

The U.S. Department of Agriculture monitors the food security of the Nation's households—the extent to which they can consistently access adequate food for active, healthy living—using data from an annual, nationally representative household survey. That survey, the Current Population Survey Food Security Supplement (CPS-FSS), is sponsored by the Economic Research Service (ERS) and conducted by the U.S. Census Bureau as an annual supplement to its monthly Current Population Survey. The questions used to assess households' food security elicit information about experiences and behaviors that typically occur when households are having difficulty meeting their food needs (Hamilton et al., 1997a; Hamilton et al., 1997b). The questions differ in the severity of the conditions they reference, from worrying that food would run out to children not eating for a whole day. Each question specifies that the behavior or experience in question occurred because there was not enough money for food.

Responses to 10 questions (18 if there are children in the household) are combined in a household food security scale in order to assess the level of severity of food insecurity in each household (Hamilton et al., 1997b; Bickel et al., 2000). Households are classified as to food security status—high, marginal, low, and very low—based on the number of food-insecure conditions they report. Item assessment and selection were based on the single-parameter logistic item response theory (IRT) (Rasch) measurement model to ensure that such raw-score-based classification was justified (Hamilton et al., 1997b). Rasch model-based analysis was also used to identify comparable levels of severity on the 18-item and 10-item scales, to monitor the stability of the measure over time and in various other surveys, and to assess comparability of the measure across demographic and linguistic groups.

In 2003-06, an expert panel convened by the Committee on National Statistics (CNSTAT) of the National Academies conducted a thorough review of the food security measurement methods. ERS requested the review by CNSTAT to ensure that the measurement methods USDA uses to assess households' access—and lack of access—to adequate food and the language used to describe those conditions are conceptually and operationally sound and that they convey useful and relevant information to policy officials and the public. The panel, which included economists, sociologists, nutritionists, statisticians, and other researchers, recommended that USDA continue to measure and monitor food insecurity regularly in a household survey, affirmed the appropriateness of the general methodology currently used by USDA, and suggested several ways in which the methodology might be refined (NRC, 2006).

USDA made several of the recommended changes that did not require confirmatory research:

- In 2006 (in the report on the 2005 data), USDA began using new labels to describe ranges of severity of food insecurity, removing the word "hunger" from the label for the more severe range.

- In the December 2006 survey, USDA changed the “resource constraint” wording of several of the food security questions to standardize the wording across questions.
- In the December 2007 survey, USDA reordered the food security questions so that all the child-referenced questions are grouped together following the adult-referenced questions.
- ERS commissioned two studies of the duration or recurrence of food insecurity over 5 or more years (Wilde et al., 2011; Ryu and Bartfeld, 2011).

The CNSTAT panel also recommended that USDA consider several technical enhancements to the statistical methodology used to assess items and scales and to classify households as to food security status. This report comprises findings from assessments of five of the most salient potential technical enhancements recommended by the panel:

- Incorporate frequency-of-occurrence information of all items for which it is available into the main measure using polytomous IRT models rather than the current dichotomous model.
- Treat items with frequency followup questions appropriately, for example, as a single ordered polytomous item, rather than as two independent questions.
- Allow item discrimination parameters to differ from item to item when indicated by the relevant data.
- Fit models that allow for different item parameters for households with and without children…in order to study the possibility and effects of differential item functioning.
- Develop a new classification system that reflects the measurement error inherent in latent variable models. This can be accomplished by classifying households probabilistically along the latent scale.

The initial research was conducted by Mark Nord on each topic. Amy Froelich, a statistician at the Iowa State University, who specializes in the relevant IRT statistical methods, reviewed the five research papers. The papers were revised based on further research as recommended by Froelich. The revised papers were then reviewed by Matthew Johnson, a nationally recognized expert on these measurement methods at the Columbia University Teachers College. Based on Johnson‘s recommendations, the papers were further revised to become Chapters 1-5 in this report, and Chapter 6 was added, examining effects of three of the potential enhancements in combination.

Chapter 1

Assessing a Polytomous Rasch Model To Include the Frequency of Occurrence of Food-Insecure Conditions

Abstract

Measures of food security calculated from polytomous-coded items incorporating frequency-of-occurrence information collected in the Current Population Survey-Food Security Supplement (CPS-FSS) are compared with measures calculated from dichotomous-coded items that omit that information. The extent of measurement error for the two models is compared. The extent to which the greatest severity of food insecurity experienced during a year and the greatest severity of food insecurity experienced frequently or chronically during the year represent distinct latent traits is explored. Findings indicate that either a polytomous or dichotomous model fit the response data reasonably well. However, reductions in measurement error of prevalence rates routinely monitored by USDA that would result from adoption of a polytomous model would be small—perhaps near zero. Furthermore, there is some evidence that measures of "ever during the year" and "frequent/chronic" food insecurity represent distinct dimensions of food insecurity, and that these two different temporal patterns of food insecurity differentially affect households with different economic and demographic characteristics. The findings suggest that it may not be appropriate to represent the two dimensions with a single polytomous scale and that, in any case, doing so would provide small to minimal gains in measurement precision.

Background

The Committee on National Statistics (CNSTAT) panel that reviewed USDA's food security measurement methods recommended that USDA consider several "more flexible alternatives to the dichotomous Rasch model that underlies the current food insecurity classification scheme" (NRC, 2006). One such alternative identified by the panel was, "Modeling ordered polytomous item responses by ordered polytomous [items] rather than dichotomized item response functions."

The current standard U.S. measures of household food security are calculated almost entirely from dichotomous items that indicate whether a condition or behavior occurred at any time during a stated reference period—usually the previous 12 months (Hamilton et al., 1997b; Bickel et al., 2000). The CPS-FSS, the primary data source used by USDA to monitor the Nation's food security, collects additional information about how often the conditions or behaviors occurred during the reference period, but with two exceptions (three for households with children), the frequency-of-occurrence information is not incorporated into the measure.[1] The CNSTAT recommendation is to incorporate this additional information into the official measure by using multiple-category (polytomous) indicator items.

[1]None of the other national-level surveys that include the household food security module collect the extensive frequency-of-occurrence information collected by the CPS-FSS.

It seems likely that providing information about how often food-insecure conditions occur would add value to the measurement of food insecurity. The current method correctly identifies households that were food secure throughout the year (subject to households reporting accurately). However, the measure provides less satisfactory information about households that had low or very low food security at some time during the year because it lacks detail on how frequently the conditions occurred. Additional information on the frequency and duration of food insecure episodes at various levels of severity may be important for assessing the severity of these conditions and for designing and assessing policies and programs to alleviate them.

Incorporating frequency-of-occurrence information into the official measure by way of a polytomous Rasch model is one way to provide this information, provided the data adequately meet assumptions of that model. But a polytomous model is not the only way to accomplish that objective, and, depending on theoretical, empirical, and communication issues, such a model may not be the preferred method. Nord and Radimer (2005) suggest an alternative method that would use two scales, one representing severity of insecurity (the most severe condition that occurred during the reference period) and second representing the frequency of insecurity (the most severe condition that occurred frequently or chronically during the reference period). Cross-classification of households by the two measures then provides a more complete representation of temporal patterns of food insecurity.

An important consideration is whether, and to what extent, severity of insecurity (the most severe condition that occurred during the reference period) and frequency of insecurity constitute a single dimension or two distinct dimensions. The two alternatives have different implications for the appropriateness and potential contribution of a polytomous model:

- If severity and frequency (as reported) represent a single dimension, then a polytomous model may fit the data. The statistical advantage of the polytomous model, however, would only be added precision. It would not improve the measure's representation of frequency or duration of food insecurity, because under the assumption of unidimensionality, each level of severity implies a specific level of frequency. In this case, a decision to use or not use a polytomous model would depend to a great extent on the practical consideration of how readily results are communicated to a policy and public audience. The measure calculated from dichotomous items is relatively easy to explain. Households with a given raw score can readily be characterized as having reported specific conditions and denied others. Such characterization may be more difficult in the case of a polytomous measure. Whether the gain in measurement precision offsets the loss of communicability is primarily a pragmatic issue.
- If severity and frequency represent substantially different dimensions, then a polytomous Rasch model will not be appropriate. The statistical sufficiency (and perhaps ordinality) of the polytomous raw score cannot be assured in the absence of conditional independence of items. The upper (more severe) thresholds will be positively correlated across items, even among households with the same total raw score. In this case, for example, a household with a high severity but low frequency is not a statistical outlier, but rather a distinctive type on a two-dimensional construct. Such a reality can only be adequately represented by two separate measures, or by a cross-classification based on them.

Nord and Radimer (2005) present evidence that severity of food insecurity (greatest severity of food insecurity experienced at any time during the year) is a substantially different dimension than frequency of food insecurity (greatest severity of food insecurity experienced frequently or chronically during the year). If this assessment is correct, then the two-measure approach is preferred both statistically and for communicating readily to policy officials and the public.

In the present study, the CPS-FSS food security response data are fit to a polytomous Rasch model, incorporating the full range of frequency-of-occurrence responses available in the data. Prior to revisiting the issue of possible bidimensionality, it seems worthwhile to assess whether the data meet the first-order condition of adequate fit to the polytomous model. The two dichotomous-item models proposed by Nord and Radimer (2005) are also estimated, and the respective scales are evaluated for each household. Next the extent to which the polytomous model would reduce measurement error compared with the current standard model is assessed.

Finally, residual inter-item correlations of response in the highest frequency-of-occurrence category and correlations between the two dichotomous scales are examined to assess the extent of bidimensionality. The extent to which the two dimensions are differentially associated with household characteristics typically used as breakouts in USDA reports is explored in multivariate logistic regression models.

Data and Methods

Data

The main analyses used nationally representative survey data from the CPS-FSS for the years 2003, 2004, and 2005 for households with incomes less than 185 percent of the poverty line. Most households with incomes higher than 185 percent of the poverty line were screened out of the food security questions; omitting those households avoids spurious associations of food insecurity with income that could result from the screen (since the screen depends partly on income), Analyses were limited to the adult and general household items in the module to avoid complications associated with the bidimensionality of adult and child items (Nord and Bickel, 2002). Households with missing responses to one or more food security questions or to any of the frequency-of-occurrence followup questions (a small proportion of households) were omitted, resulting in a basic analysis sample of 34,911 households.

Data from the December 2010 CPS-FSS for households with incomes less than 185 percent of the poverty line were used to assess the extent to which measurement error in prevalence rates would be reduced by using a polytomous model rather than the current standard methods. The 2010 data were used for this analysis because 2010 was the first year for which the Census Bureau calculated replicate weights for the Food Security Supplement to support calculation of sampling errors using balanced repeated replication (BRR) methods. Errors calculated using BRR methods are, inherently, a combination of sampling error and measurement error. In the analyses presented here, sampling error is the same for the two measurement methods, since they are based on the same data, so differences in BRR-estimated errors can be ascribed to differences in measurement error.

Food Security Measurement Models

There are eight adult-referenced and general household questions in the food security module (see appendix A). In the CPS-FSS, frequency of occurrence of food-insecure conditions is collected using two different methods:

- Method 1: A condition is described, and the respondent is asked whether this was often, sometimes, or never true for his or her household during the past 12 months. This method is used for the first three questions.
- Method 2: Respondents are first asked whether a specific behavior or condition ever occurred during the previous 12 months. If they answer "yes," they are then asked, "How often did this happen—almost every month, some months but not every month, or in only 1 or 2 months?"

No frequency-of-occurrence information is elicited about question 7 (about losing weight because of insufficient money for food).

Five scales based on these data were calculated:

- Polytomous scale: This full-information scale uses all of the response categories in the 12-month adult and household questions. Method 1 responses were coded 0 for "never," 1 for "sometimes," and 2 for "often." Method 2 responses were coded 0 for "no," 1 for "yes, in only 1 or 2 months," 2 for "yes, in some months but not every month," and 3 for "yes, in almost every month." Question 7 was coded 0 for "no" and 1 for "yes."
- "Ever during the year" scale: Method 1 responses were coded 0 for "never" and 1 for "sometimes" or "often." Method 2 responses were coded 0 for "no" and 1 for "yes" (without reference to the frequency-of-occurrence followup). Question 7 was coded 0 for "no" and 1 for "yes."
- "Frequent/chronic" scale: This is the frequent/chronic scale proposed by Nord and Radimer (2005). Method 1 responses were coded 0 for "never" or "sometimes" and 1 for "often." Method 2 responses were coded 0 for "no," "yes, in only 1 or 2 months," and "yes, in some months but not every month," and 1 for "yes, in almost every month. Question 7 was omitted from this scale.
- Ever + frequent/chronic trichotomous scale: This partial-information polytomous scale comprises the items in the ever-during-the-year and the frequent/chronic scales. Alternatively, it can be thought of as derived from the full-information polytomous scale by collapsing the two intermediate categories in the four-level items, "yes, in only 1 or 2 months," and "yes, in some months but not every month."
- Standard adult food security scale: This is the scale used by USDA for official food security statistics for households without children. It is also used for many research purposes for all households, with or without children. Two items, *Adult cut size of meals or skipped meals and Adult did not eat whole day* were coded as trichotomies, 0 for "no," 1 for "yes, in only 1 or 2 months," and 2 for "yes, in some months but not every month"

or "yes, in almost every month." The remaining 6 items were coded as ever-versus-never dichotomies. The standard scale was used only in comparisons to assess the extent to which the polytomous scale would reduce measurement error. Assessment of this scale and, in particular, of the two frequency-of-occurrence questions included in it, is the topic of Chapter 2. For analytic purposes of the present study, it was important to separate as completely as possible the frequency of food insecurity from the most severe level experienced during the year, so the two frequency-of-occurrence items in the standard scale were omitted from the ever-during-the-year scale.

Household Demographic and Economic Characteristics

Variables representing annual income, employment of adults in the household, education of the most highly educated adult, household composition, race and Hispanic ethnicity, citizenship, home ownership, "recent move" status, metropolitan status of residence, and region of residence were calculated from data elements collected in the core labor-force section of the Current Population Survey.

Assessing Model Fit

Psychometric characteristics of the first three scales (the polytomous scale, the ever-during-the-year scale, and the frequent/chronic scale) were assessed using standard Rasch model-based statistical methods. The polytomous model was estimated as a partial credit Rasch model[2] using joint (or unconditional) maximum likelihood (JML) implemented in Winsteps software. The two dichotomous-item models were estimated using conditional maximum likelihood (CML) methods implemented by software developed by ERS.[3] Estimates of the level of household (respondent) severity corresponding to each raw score in these three models were based on maximum likelihood (ML) methods given the item parameters. Mean-square item infit and outfit statistics were examined to assess the fit of the data to each measurement model. For the polytomous model, threshold-specific item-fit statistics were also calculated, using only responses in the category immediately above the threshold and the category immediately below the threshold. Expected probabilities were calculated, conditional on the response being in one of those two categories.

After omitting households with extreme responses (households that denied all items in a scale or with maximum raw scores), sample sizes for estimating the models were as follows: 15,178 for the polytomous model, 14,607 cases for the ever-during-the-year scale calculated from dichotomized items, and 5,237 cases for the Nord-Radimer frequent/chronic scale calculated from dichotomized items.

Assessing the Extent to Which the Full-Information Polytomous Scale Reduces Measurement Error

Standard errors were calculated for prevalence rates of food insecurity and very low food security based on the standard adult scale and the full-information polytomous scale using balanced repeated replication (BRR methods). Errors were compared between the two measures for all households with

[2]The partial-credit polytomous Rasch model allows the inter-threshold distance (in logits) to differ between items.

[3]The ERS software, ERS Rasch, implements CML estimation methods based on equations by Fischer and Molenaar (1995) and has been tested using simulated data and by comparing to Winmira and other commercially available software applied to actual data.

incomes less than 185 percent of the poverty line and for selected low-income demographic and geographic subpopulations. Food security status on each scale was assigned based on raw score, with thresholds on the polytomous scale selected to most nearly replicate the national-level prevalence rates based on the standard scale.

The BRR methods were implemented in the SAS SurveyFreq procedure using replicate weights provided by the Census Bureau for the December 2010 CPS-FSS.[4]

Error estimates calculated by BRR inherently combine sampling error and measurement error. Indeed, the two sources of error cannot be differentiated using BRR methods. These analyses compared errors on prevalence estimates from the two measures calculated from the same samples, so sampling error was the same for the two measurement methods, and differences in estimated errors represented differences in measurement error.

Assessing Dimensionality

Conditional independence of items in a Rasch model is typically examined by a principal-components factor analysis of standardized residuals after fitting the data to a Rasch model. In this case, however, the concern is that responses in categories indicating more frequent occurrence might be correlated across some or all variables, rather than overall response to the variables. For example, households with occasional, but severe, spells of food security, might generally have many "sometimes" responses, but few "often" responses. Other households with the same raw score, but with chronic food security might generally affirm fewer items but at higher levels of frequency. Such correlations across responses would not be detected by a standard factor analysis of residuals.

Cross-level conditional dependence using one direct method and one indirect method was assessed. The direct method began by estimating the ever + frequent/chronic trichotomous model under Rasch partial-credit assumptions using CML methods.[5] Based on item parameters estimated from that model, two sets of residual correlations between each pair of items were calculated and compared. The first set was based on "ever" versus "never" responses. The second was based on response in the most severe response category (i.e., "often" or "almost every month") versus response in the mid-level category (i.e., "sometimes," "yes, in only 1 or 2 months," or "yes, some months but not every month"). Responses of "no" or "never" were omitted from calculation of the second set of residual correlations. Responses were also omitted if the raw score constrained any combination of responses to two items. (For example, responses with raw score 3 were not used to calculate the second set of residual correlations because it would be impossible for both items to be affirmed in the "often" or "almost every month" category.)

Under partial-credit Rasch model assumptions, both types of correlations should be zero. In actual data that only approximately meet Rasch assumptions, there will be some residual correlations. If frequency of occurrence is a distinct dimension from highest severity during the year, the second type of residual correlation (frequent versus sometimes) will be more positive than the first type (ever versus never) for most or all item-pairs.[6]

[4]Replicate weights were first provided for the CPS-FSS in 2010 survey. The weights are based on the Fay method and, following specifications provided by the Bureau, the option "Fay=.5" was specified in the SAS SurveyFreq procedure.

[5]This assessment was conducted based on the trichotomous model rather than the full-information polytomous model because of a software limitation. My CML software can estimate trichotomous items, but not 4-category items. CML estimation is essential for this analysis because estimates of residual correlations based on JML methods are not consistent, and the extent of bias is not known for the inter-level correlations of interest here.

[6]The software was tested on simulated trichotomous data that were stochastically Rasch-consistent, and that had item parameters and distribution across nonextreme raw scores similar to those in the CPS-FSS data for 2003-05. All residual correlations of both types were zero or very near zero in these simulated data (none exceeded .02 in absolute value).

The indirect method for assessing cross-level conditional dependence examined the correlation between household measures based on the ever-during-the-year scale and the frequent/chronic scale, limiting the analysis to households that were nonextreme on the frequent/chronic scale. Under Rasch-model assumptions of conditional independence, these measures would be nearly perfectly correlated if the scales had large numbers of well-conditioned items. Perfect correlation cannot be expected in the case of the food security scales because of the small number of items. The observed correlation was compared with the corresponding correlation calculated from simulated data that were stochastically consistent with polytomous Rasch-model assumptions and were similar in relevant characteristics to the CPS-FSS data. Specifically, the distribution of true food insecurity across the simulated households and the item parameters used to generate the simulated response data were selected so that the resulting distribution of households across raw score groups was similar to that in the CPS-FSS data, and item parameters estimated from the simulated data were similar to those estimated from the CPS-FSS data. An interscale correlation substantially lower in the CPS-FSS data than in the simulated data would suggest that frequency of occurrence represents a substantial second dimension in the data and violates the Rasch model assumption of conditional independence.[7] The simulation was replicated 1,000 times to provide an estimate of the expected variance of the interscale correlation if the items are, in fact, conditionally independent.

[7]This inference is valid only if discrimination of the two scales is equal. If one scale discriminates more poorly than the other, correlation between the two scales will be lower. In the data analyzed here, item-fit statistics in the polytomous model suggest that discrimination in the two dichotomous scales should be similar.

Two approaches were used to assess the practical importance of a second dimension suggested by the analyses just described, First, differences in the associations of the two dichotomous-item scales with households' demographic and economic characteristics were examined using logistic regression analyses. Following methods described in Nord and Radimer (2005), frequent or chronic food insecurity, as measured by the frequent/chronic scale, was regressed on selected household demographic and economic characteristics, controlling for the level of severity as measured by the ever-during-the-year scale. Substantial and statistically significant coefficients on household characteristics would indicate that severity and frequency of food insecurity represent dimensions that differ sufficiently to be of practical importance.

The second approach contrasted characteristics of two types of households based on a cross-tabulation of the ever-in-the-year scale and the frequent-chronic scale:

- Those with severe, but not frequent food insecurity. These households had raw scores of 6-8 on the ever-during-the-year scale, indicating very low food security at some time during the year, but raw scores of 0-2 on the frequent/chronic scale, indicating that they were not frequently or chronically food insecure even in the low food security range.
- Those with frequent, but not severe, food insecurity. These households had raw scores of 3-5 on both scales, indicating that they had frequent low food security, but did not experience very low food security at any time during the year.

A regression analysis was conducted with the analysis sample limited to households in those two cells. Severe-but-not-frequent food insecurity status was regressed on selected household demographic and economic characteristics. All households had raw scores of 6 to 10 on the polytomous scale comprising the ever-in-the-year and frequent/chronic items. A set of dummy variables was included in the regression model to control for overall severity on the combined scales. Substantial and statistically significant coefficients on household characteristics would indicate that different types of households experience these two temporally distinct conditions in spite of having the same score on the polytomous scale.

Findings

Polytomous Rasch Model

The response data generally meet the Rasch-model assumption of equal discrimination to an acceptable degree. Summary infit statistics for each item ranged from 0.78 to 1.24 (table 1-1). Infit statistics for *Could not afford to eat balanced meals* and *Adult did not eat whole day* were slightly elevated. However, the former immediately precedes a screen, which is known to bias fit statistics upward modestly, and the latter is the most severe item, for which fit statistics are known to be biased upward (modestly for infit and substantially for outfit) when item parameters are estimated in JML (Nord, 2006). Threshold-specific infit statistics were all acceptably near unity. Item outfit statistics were elevated for several items, but followed an expected pattern given the screeners implemented in the module and the distortions due to JML estimation.

Ever-During-the-Year Dichotomous-Item Model

The CPS-FSS data also fit a dichotomous Rasch model reasonably well when coded into items with 1 indicating any affirmative response (table 1-2). Item infit statistics are quite good except for *Could not afford to eat balanced meals* (1.29). The infit of this item is biased upward by the screening imposed at administration of the food security series (Nord, 2006). Households that say “never” to the first three items are skipped over the remaining items (unless they report that they sometimes or often did not have enough to eat in response to the food sufficiency question, which is not part of the scale) and negative responses are imputed. When the effect of this screen is obviated by omitting households that denied the first (least severe) item, and omitting that item from the scale, the infit of *Could not afford to eat balanced meals* was just 1.15 (analysis not shown). The infit was also reasonably low (1.17, analysis not shown) when the model was estimated from the subsample of households that were screened into the second block of questions based on the food sufficiency question (responses of sometimes or often did not have enough to eat), which allowed inclusion of the least severe scale item. The outfit statistic for *Could not afford to eat balanced meals* (2.47 in the main model) however, remained somewhat high, even with screening effects obviated. Household severity parameters for this “ever during the year” scale are provided in table 1-3.

Frequent/Chronic Dichotomous-Item Model

The data also fit a dichotomous Rasch model well when coded affirmative only for the response indicating maximum frequency of occurrence (“often”

Table 1-1

Item severity parameters and fit statistics for polytomous partial-credit Rasch food security model using all available frequency-of-occurrence information

Item[1] and threshold	Severity		Item-fit	
	Measure	50-percent probability[2]	Infit	Outfit
Worried food would run out			1.13	1.14
T1: Sometimes vs. never	5.54	5.53	1.10	2.69
T2: Often vs. sometimes	10.10	10.11	1.10	1.17
Food bought did not last			.96	.95
T1: Sometimes vs. never	6.93	6.91	.91	1.11
T2: Often vs. sometimes	10.80	10.82	1.01	.92
Could not afford to eat balanced meals			1.24	1.27
T1: Sometimes vs. never	7.52	7.48	1.19	1.51
T2: Often vs. sometimes	10.66	10.70	1.13	1.49
Adult cut size of meals or skipped meals			.83	.66
T1: In 1 or 2 months vs. never	10.16	9.15	.98	.77
T2: Some months but not every month vs. 1 or 2 months	8.80	9.56	.92	.90
T3: Almost every month vs. some months but not every month	10.72	10.89	.91	.89
Respondent ate less than he/she felt he/she should			.78	.54
T1: In 1 or 2 months vs. never	10.12	9.15	.98	.69
T2: Some months but not every month vs. 1 or 2 months	8.81	9.57	.89	.85
T3: Almost every month vs. some months but not every month	10.90	11.04	.86	.83
Respondent hungry but did not eat			.95	.60
T1: In 1 or 2 months vs. never	11.68	10.40	1.02	.66
T2: Some months but not every month vs. 1 or 2 months	9.71	10.68	.91	.86
T3: Almost every month vs. some months but not every month	11.60	11.80	.89	.86
Respondent lost weight (yes vs. no)	11.40	11.40	.89	.48
Adult did not eat whole day			1.22	.54
T1: In 1 or 2 months vs. never	12.92	11.42	1.01	.54
T2: Some months but not every month vs. 1 or 2 months	10.45	11.65	.95	.95
T3: Almost every month vs. some months but not every month	12.60	12.79	1.11	1.29

[1]The full wording of each question includes explicit reference to resource limitation, e.g., "…because there wasn't enough money for food."
[2]The level of severity at which the probability of response in any category above the threshold is .5.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2003, 2004, 2005 for households with incomes less than 185 percent of the Federal poverty line and with complete data on the eight household and adult food security items and followup frequency-of-occurrence items (N=15,178, omitting extreme responses).

or "almost every month"). Item infit statistics ranged from 0.84 to 1.19, well within an acceptable range (table 1-4). Item outfit statistics were also quite good except for "Could not afford to eat balanced meals," which was somewhat high. Household severity parameters for this "frequent/chronic" scale are provided in table 1-5.

Measurement Error

The principal purported advantage of the polytomous model is that it would, in theory, reduce measurement error because it makes use of more information. It appears, however, that the reduction in measurement error would be small for very low food security and small or zero for food insecurity. At the national level, combined sampling and measurement error for the polytomous measure was slightly larger than that of the current standard measure for the

Table 1-2

Item severity parameters and fit statistics from dichotomous Rasch model measuring the most severe condition of food insecurity that occurred at any time during the year

Item[1]	Severity parameter	Item-fit Infit	Item-fit Outfit
Worried food would run out	3.62	1.00	2.87
Food bought did not last	4.66	.90	1.57
Could not afford to eat balanced meals	5.14	1.29	2.47
Adult cut size of meals or skipped meals	6.95	.82	.73
Respondent ate less than he/she felt he/she should	6.93	.78	.63
Respondent hungry but did not eat	8.77	.88	.62
Respondent lost weight	9.62	.98	.71
Adult did not eat whole day	10.31	.98	.53

[1]The full wording of each question includes explicit reference to resource limitation, e.g., "… because there wasn't enough money for food."

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2003, 2004, 2005 for households with incomes less than 185 percent of the Federal poverty line and complete data on the eight household and adult food security items (N=14,607 omitting extreme responses).

Table 1-3

Household severity measure by raw score on ever-during-the-year measure of food insecurity, based on item parameters as presented in table 1-2

Raw score	Severity measure
1	3.59
2	4.85
3	5.93
4	6.98
5	8.05
6	9.17
7	10.44
8[1]	11.38

[1]Evaluated at 7.5.

Source: USDA, Economic Research Service analysis.

Table 1-4

Item severity parameters and fit statistics from dichotomous Rasch model measuring the most severe condition of food insecurity that occurred frequently or chronically during the year

Item[1]	Severity parameter	Item-fit Infit	Item-fit Outfit
Worried food would run out	5.28	1.08	1.29
Food bought did not last	6.18	.97	.97
Could not afford to eat balanced meals	6.10	1.19	1.40
Adult cut size of meals or skipped meals	6.75	.89	.87
Respondent ate less than he/she felt he/she should	6.92	.84	.77
Respondent hungry but did not eat	8.19	.86	.64
Adult did not eat whole day	9.58	1.05	.95

[1]The full wording of each question includes explicit reference to resource limitation, e.g., "… because there wasn't enough money for food."

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2003, 2004, 2005 for households with incomes less than 185 percent of the Federal poverty line and with complete data on the eight household and adult food security items and followup frequency-of-occurrence items (N=17,791 omitting extreme responses).

Table 1-5

Household severity measure by raw score on frequent/chronic measure of food insecurity, based on item parameters as presented in table 1-4

Raw score	Severity measure
1	4.71
2	5.72
3	6.51
4	7.29
5	8.18
6	9.41
7[1]	10.38

[1]Evaluated at 6.5.

Source: USDA, Economic Research Service analysis.

prevalence of food insecurity and about 6 percent smaller than that of the current standard measure for very low food security (fig. 1-1). The differences between the measures are smaller when they are adjusted for differences in prevalence rates. The discrete thresholds are not quite perfectly comparable, and larger prevalence rates generally have larger sampling errors (until the prevalence exceeds 50 percent). At the national level, adjusting for the differences in prevalence rates would lower the ratio for food insecurity shown in figure 1 by about 3 percent and raise the ratio for very low food security by about 5 percent, leaving both very near unity.[8]

The gains in measurement precision due to the polytomous model are small, in part because of the relatively high severity of the upper thresholds of the polytomous items. The threshold for food security on the metric of the scale in table 1-1 is about 7.5, and the threshold for very low food security is about 10.0. Only two of the T2 and T3 item thresholds (based on the Rasch-Thurstone 50-percent probability transformations) are below 10, and one of those two is already included in the standard measure. Adding the information from these rather severe item components to the scale reduces measurement error primarily in the high-severity end of the scale, relatively far above the threshold for the highest severity level monitored by USDA.

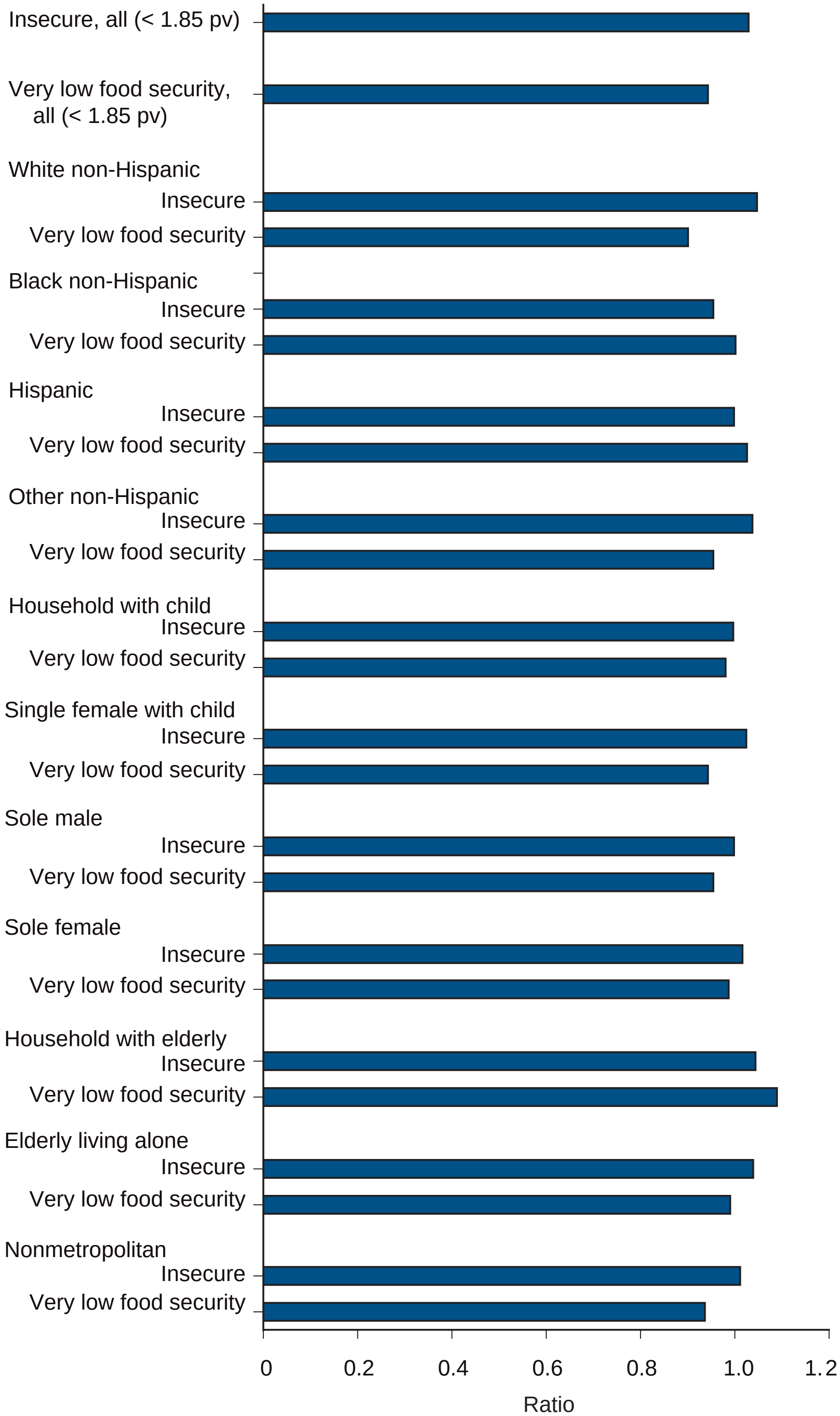
An upper limit of the extent to which the polytomous scale could improve classification is given by the proportion of households classified differently by the polytomous scale and the current standard scale. Direct cross-classification cannot provide this proportion because discrete thresholds on the two scales are not precisely equivalent in terms of prevalence rates. Interpolation across nearly equivalent thresholds (analysis not shown) suggests that at equivalent thresholds, for the low-income population analyzed throughout this chapter, about 3 percent of households would be classified differently at the food insecure threshold (1.5 percent in each direction), and about 2.4 percent would be classified differently at the very low food security threshold (1.2 percent in each direction). Because prevalence rates were about 25 percent for food insecurity and 10 percent for very low food security, these would be nontrivial improvements if the polytomous scale could be taken as the gold standard. However, the small extent of reduction in measurement error reported above, as well as findings in the rest of this chapter, suggest that some of the difference in classification results from problems inherent in the polytomous scale rather than lack of precision of the standard scale.

[8]Sampling error for a prevalence rate is proportional to the square root of PQ, where P is the prevalence and Q is its complement. This proportionality was used to adjust the ratio of BRR errors for the two measures.

Prevalence estimate errors between the full-information polytomous scale and the current standard scale were compared with measurement error taken into account for both scales. That is, the probability of food insecurity (/very low food security) was calculated for each raw score. The mean of household probabilities for the population and for subpopulations were taken as the estimated prevalence rates. Results of that analysis are reported in chapters 5 and 6. In short, the probabilistic method reduced measurement error for both the polytomous and current standard method, but differences in errors between the two measures, when both used the probabilistic method, were small to negligible, similar to the results presented in this chapter.

Figure 1-1

Ratio of combined sampling and measurement error of prevalence rates based on polytomous model to error based standard 10-item model



Notes: Standard errors were calculated by balanced repeated replication using replicate weights provided by the U.S. Census Bureau as part of the public-use data.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplement of December 2010 for households with incomes less than 185 percent of the Federal poverty line.

Assessing Whether Severity and Frequency Represent Distinct Dimensions

The level-specific residual correlations indicate that frequency of occurrence represents a distinct dimension from the greatest severity that is experienced during the year. With the exception of one item-pair, residual correlations at the upper threshold were more positive than at the lower threshold (fig. 1-2). For about two-thirds of the item-pairs, residual correlations were positive at the upper threshold and negative or near zero at the lower. For item-pairs for which residual correlations were positive at both thresholds, the correlation at the upper threshold was generally two to three times that at the lower threshold. Residual correlations at the upper threshold exceeded .4 for five item-pairs. The sample distribution of this statistic was not calculated, but the size and consistency of the differences leaves little doubt that frequency of occurrence does, indeed, constitute a second dimension.

The Pearson correlation between household severity measures based on the two dichotomous-item scales (i.e., with household measures as in tables 1-3 and 1-5) also suggests a nontrivial bi-dimensionality in the response data. Among households with nonextreme responses on both scales, the Pearson correlation between the household severity measures based on the ever-during-the-year scale and the frequent/chronic scale was .56 (analysis not shown). The expected correlation based on simulated data that are stochastically consistent with Rasch model assumptions, and otherwise similar to the CPS-FSS data, is .64. A Monte Carlo replication with 1,000 replicates, each of approximately the same number of cases as in the CPS-FSS data, indicated that the difference in correlations was highly significant (the standard error was .006). The lower than expected correlation between the scales suggests that responses indicating higher frequency of occurrence do not indicate simply a more severe condition on a single latent trait, but a somewhat different condition highly correlated with that trait yet distinct from it.

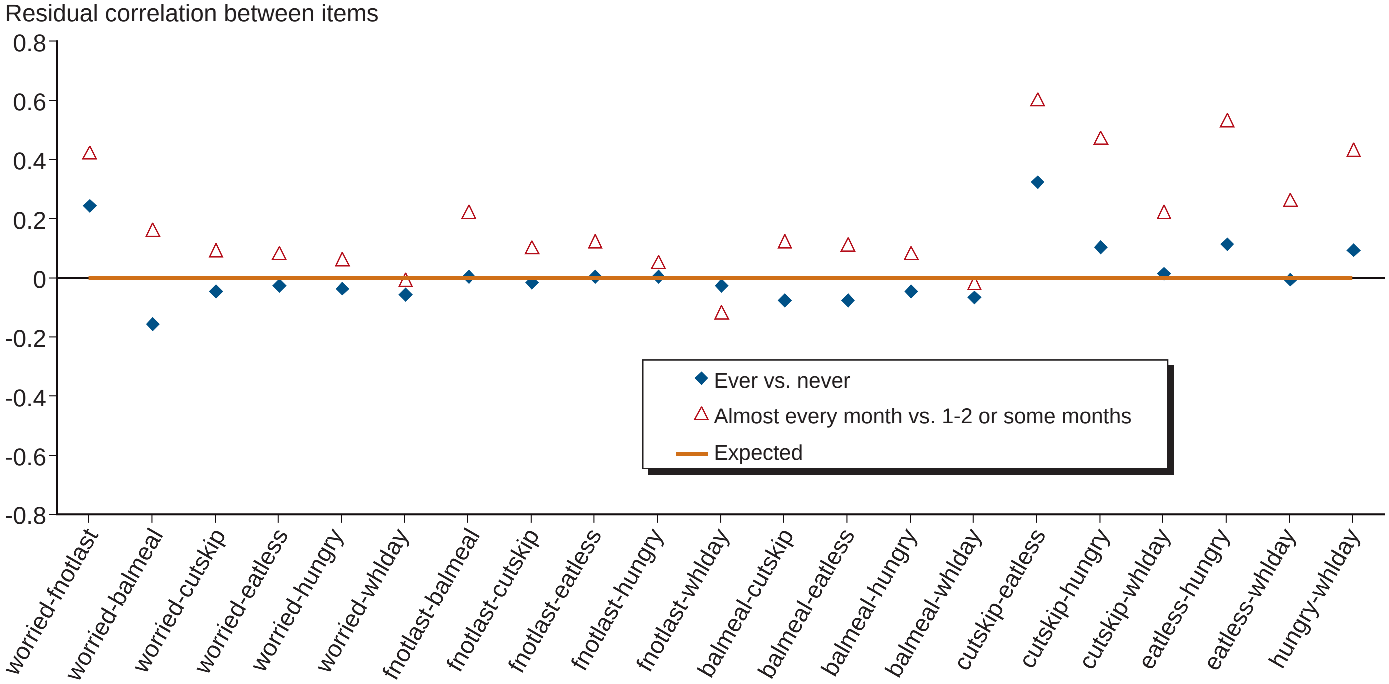
A possible threat to this interpretation is that the correlation between these scales would also be reduced if discrimination of one of the scales were substantially lower than the other. However, there is little evidence of substantial difference in discrimination between the items in the two scales. Although infit statistics were somewhat higher at the upper (T2) threshold than at the lower (T1) for three items, differences were in the opposite direction for two items and near zero for the remaining two items (table 1-6). Taken together, these do not suggest sizeable differences in discrimination between the two sets of items.

Determinants of Frequent/Chronic and Ever-During-the-Year Food Insecurity

The practical implications of the bidimensionality in the response data described above are suggested by the associations of the different measures of food insecurity with household demographic and economic characteristics. Households that experienced frequent or chronic food insecurity had different economic and demographic characteristics than those that experienced more severe food-insecure conditions but of shorter duration.

Figure 1-2

Inter-item residual correlations based on ever-versus-never and frequent/chronic versus sometimes responses



Note on horizontal axis labels:
Worried-fnotlast=conditional correlations between the items, “Worried food would run out” and “Food bought did not last.”

Labels on X axis refer to:
worried=Worried food would run out.
fnotlast=Food bought did not last.
balmeal=Could not afford to eat balanced meals.
cutskip=Adult cut size of meals or skipped meals.
eatless=Adult ate less than he/she felt he/she should.
hungry=Respondent hungry, but did not eat.
whlday=Adult did not eat whole day.

Notes: Correlations are from a single model with items coded as trichotomies. Often/sometimes/never items were coded as such; yes/no-with-follow-up items were coded never/in only 1-2 months or some months but not every month/almost every month. Inter-item residual correlations were calculated by comparing the observed dichotomous cross-tabulation of responses in the two affirmative response levels for each pair of items with the expected cross-tabulation based on conditional maximum likelihood-based probabilities for all possible response patterns.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2003, 2004, 2005 for households with incomes less than 185 percent of the Federal poverty line and with complete data on the eight household and adult food security items.

Table 1-6

Item severity parameters and fit statistics from trichotomous partial-credit Rasch model combining items from the ever-during-the-year and frequent/chronic food security scales

	Severity		Item-fit	
Item[1] and threshold	Measure	50-percent probability[2]	Infit	Outfit
Worried food would run out			1.10	1.11
T1: Sometimes vs. never	3.24	3.22	1.02	1.91
T2: Often vs. sometimes	6.92	6.95	1.11	1.30
Food bought did not last			.94	.93
T1: Sometimes vs. never	4.21	4.17	.88	1.00
T2: Often vs. sometimes	7.60	7.63	1.01	1.06
Could not afford to eat balanced meals			1.27	1.32
T1: Sometimes vs. never	4.68	4.62	1.19	1.56
T2: Often vs. sometimes	7.45	7.51	1.18	2.60
Adult cut size of meals or skipped meals			.86	.80
T1: In 1 or 2 months or some months vs. never	6.32	6.12	.85	.79
T2: Almost every month vs. 1 or 2 months or some months	7.66	7.86	.90	1.03
Respondent ate less than he/she felt he/she should			.80	.75
T1: In 1 or 2 months or some months vs. never	6.30	6.11	.83	.77
T2: Almost every month vs. 1 or 2 months or some months	7.85	8.01	.84	.83
Respondent hungry but did not eat			.90	.64
T1: In 1 or 2 months or some months vs. never	7.78	7.48	.95	.68
T2: Almost every month vs. 1 or 2 months or some months	8.51	8.81	.83	.74
Respondent lost weight (yes vs. no)	8.26	8.26	.91	.56
Adult did not eat whole day			1.16	1.04
T1: In 1 or 2 months or some months vs. never	8.88	8.52	1.10	.93
T2: Almost every month vs. 1 or 2 months or some months	9.37	9.73	1.08	1.78

[1]The full wording of each question includes explicit reference to resource limitation, e.g., "…because there wasn't enough money for food."
[2]The level of severity at which the probability of response in any category above the threshold is .5.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2003, 2004, 2005 for households with incomes less than 185 percent of the Federal poverty line and with complete data on the eight household and adult food security items and followup frequency-of-occurrence items (N=14,950, omitting extreme responses).

Coefficients were statistically significant on about half of the household characteristics in the logistic regression of frequent or chronic food insecurity with controls for severity of food insecurity as measured by the ever-during-the-year scale (table 1-7). In this model, the most severe food-insecure condition experienced during the year was controlled by a set of dummy variables representing raw score on the ever-during-the-year scale. The associations were substantial for some variables. For example, the odds of frequent food insecurity in households with no adult in the labor force and at least one adult not in the labor force because of a work-limiting disability were 74 percent higher than in an otherwise similar household with a full time worker and with the same raw score on the ever-during-the-year scale (analysis not shown; odds ratio corresponds to the logistic coefficient of 0.56). The odds of frequent food insecurity in households in which the highest educated adult had a bachelor's degree were 37 percent less in an otherwise similar household, with the same raw score on the ever-during-the-year scale, in which the highest educated adult had only a high school education. The numbers of households in each category represented by the economic and demographic variables in table 1-7 are provided in table 1-8 along with weighted percentages in each category.

Table 1-7

Logistic regressions of food insecurity based on ever-during-the-year and frequent/chronic scales on household economic and demographic characteristics

Characteristic	(Model 1) Food insecurity at any time during the year[1]		(Model 2) Frequent/chronic food insecurity (with controls for severity of food security at any time during the year)[2]	
	Coefficient	Probability	Coefficient	Probability
Intercept	-1.52		-3.96	
Household income (reference = 150 to 185 percent of poverty line)				
Less than 50 percent of poverty line	0.56	<.001	0.22	0.067
50-75 percent of poverty line	.66	<.001	.20	.103
75-100 percent of poverty line	.52	<.001	.24	.042
100-150 percent of poverty line	.30	<.001	.04	.724
Household employment (reference = one or more full-time worker)				
All adults retired, not in labor force	-.32	<.001	.27	.084
Part-time worker for non-economic reasons, no full-time worker	.05	.333	.08	.541
Part-time worker wants to work full time, no full-time worker	.75	<.001	-.04	.823
Unemployed, no employed adult	.64	<.001	-.02	.875
Disabled, no adult in labor force	.76	<.001	.56	<.001
Other, no adult in labor force	.03	.573	.25	.077
Highest education level of adult (reference = high school or GED)				
Less than high school	.15	<.001	.21	.014
Some college, no 4-year degree	.02	.469	-.14	.084
Bachelor, other 4-year degree	-.41	<.001	-.46	.002
Graduate or professional degree	-.40	<.001	-.06	.804
Household structure (reference = two-parent with child)				
Female with child, no spouse	.16	<.001	.28	.014
Male with child, no spouse	.07	.331	.53	.003
Other household with child	-.24	.109	.31	.441
Two or more adults, no child	.04	.339	.37	.001
Woman living alone	-.10	.034	.36	.003
Man living alone	-.04	.463	.18	.164
Any elderly person in household	-.49	<.001	.20	.122
Race/ethnicity (reference = white non-Hispanic)				
Black non-Hispanic	.21	<.001	-.34	<.001
Hispanic	.03	.552	-.26	.017
Other	-.10	.105	.08	.575
Non-citizen (household reference person)	-.12	.011	-.10	.460
Homeowner	-.32	<.001	.12	.116
Moved since entering Current Population Survey	-.06	.187	-.06	.593
Metropolitan area residence (reference = metro not in central city)				
In central city	.00	.903	.04	.647
Metropolitan, central city residence not identified	.03	.473	.27	.007
Outside metropolitan area	-.06	.102	.20	.028

—Continued

Table 1-7

Logistic regressions of food insecurity based on ever-during-the-year and frequent/chronic scales on household economic and demographic characteristics—Continued

Characteristic	(Model 1) Food insecurity at any time during the year[1]		(Model 2) Frequent/chronic food insecurity (with controls for severity of food security at any time during the year)[2]	
	Coefficient	Probability	Coefficient	Probability
Census Region (reference = Northeast)				
Midwest	.14	.002	.10	.374
South	.10	.015	.10	.329
West	.20	<.001	.08	.500
Raw score on ever-during-the-year scale (reference=3)				
Raw score = 4	NIM		1.08	<.001
Raw score = 5	NIM		1.96	<.001
Raw score = 6	NIM		3.00	<.001
Raw score = 7	NIM		3.56	<.001
Raw score = 8	NIM		4.47	<.001
Somers' D	.38		.70	
Number of cases	34,911		8,421	

[1]Analysis sample for Model 1 included all households with incomes less than 185 percent of the poverty line and with valid responses to all food security items.

[2]Analysis sample as in Model 1 except restricted to households with raw score 3 or more on ever-during-the-year scale.

NIM = Not in model.

GED = General educational diploma (high-school equivalent).

Shaded characteristics are significantly associated (90-percent confidence) with frequent/chronic food insecurity even with controls for most severely food-insecure conditions that occurred at any time during the year.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2003, 2004, 2005 for households with incomes less than 185 percent of the Federal poverty line.

Most of the statistically significant coefficients in the frequent/chronic logistic regression analysis were in the same direction as those in the ever-during-the-year model, but there were some notable exceptions. Households in which all adults were retired were less likely to be food-insecure (at any time during the year) than otherwise similar households with a full-time worker. However, if they were food insecure at any time during the year, they were more likely to be frequently or chronically food insecure. The same was true of women living alone, while the opposite was true for households with a Black non-Hispanic reference person. Some associations with the frequent/chronic measure could result from incomplete control for food insecurity resulting from the small number of items in the ever-during-the-year scale. However, the statistically significant associations in opposite directions are not readily explained by this mechanism.

The comparison of characteristics of households in contrasting "atypical" cells in the cross-classification of households by the ever-during-the-year and frequent/chronic scales provides additional evidence that the scales measure distinct phenomena. Most households with frequent/chronic low food security experienced very low food security at some time during the year (table 1-9). Similarly, most households with very low food security at some time during

Table 1-8

Number of households (not weighted) and weighted percentage in each category in table 1-7 models

Characteristic	(Model 1)[1] Number	(Model 1)[1] Percent[3]	(Model 2)[2] Number	(Model 2)[2] Percent[3]
Household income (reference = 150 to 185 percent of poverty line)	7,508	20.67	1,195	13.13
Less than 50 percent of poverty line	5,469	16.68	1,845	22.99
50-75 percent of poverty line	4,789	14.04	1,540	18.50
75-100 percent of poverty line	5,574	16.03	1,445	17.09
100-150 percent of poverty line	11,571	32.57	2,396	28.09
Household employment (reference = one or more full-time worker)	16,611	48.85	3,781	45.24
All adults retired, not in labor force	7,738	20.48	900	10.12
Part-time worker for non-economic reasons, no full-time worker	2,792	7.47	697	7.73
Part-time worker wants to work full time, no full-time worker	787	2.44	339	4.42
Unemployed, no employed adult	1,617	5.04	693	8.64
Disabled, no adult in labor force	3,557	10.13	1,498	17.24
Other, no adult in labor force	1,809	5.60	513	6.62
Highest education level of adult (reference = high school or GED)	12,989	36.98	3,173	37.05
Less than high school	7,436	22.80	1,991	25.54
Some college, no 4-year degree	10,384	28.85	2,598	29.70
Bachelor, other 4-year degree	3,118	8.55	517	5.87
Graduate or professional degree	984	2.83	142	1.83
Household structure (reference = two-parent with child)	7,304	20.70	1,580	19.10
Female with child, no spouse	6,075	18.63	2,067	25.40
Male with child, no spouse	1,133	3.55	331	4.11
Other household with child	269	0.81	58	0.70
Two or more adults, no child	8,649	24.23	1,773	20.83
Woman living alone	7,434	19.70	1,512	16.64
Man living alone	4,317	12.38	1,100	13.23
Any elderly person in household	9,281	24.93	1,202	14.04
Race/ethnicity (reference = White non-Hispanic)	22,060	55.27	4,761	48.97
Black non-Hispanic	5,165	19.26	1,667	25.25
Hispanic	5,407	19.87	1,418	20.85
Other	2,279	5.59	575	4.93
Non-citizen (household reference person)	3,344	12.19	827	12.30
Homeowner	16,358	44.62	2,921	33.07
Moved since entering Current Population Survey	2,448	7.10	707	8.09
Metropolitan area residence (reference = metro not in central city)	8,684	29.72	2,029	28.10
In central city	8,993	31.10	2,389	34.10
Metropolitan, central city residence not identified	6,123	15.22	1,566	15.61
Outside metropolitan area	11,111	23.96	2,437	22.19
Census Region (reference = Northeast)	5,843	14.87	1,334	13.81
Midwest	8,510	21.28	1,957	21.16
South	11,924	41.74	2,953	42.11
West	8,634	22.21	2,177	22.92

[1]Analysis sample included all households with incomes less than 185 percent of the poverty line and valid responses to all food security questions.

[2]Analysis sample as in model 1 except restricted to households with raw score 3 or more on ever-during-the-year scale.

[3]Percentages within the analysis sample were calculated using sampling weights (household supplement weights).

GED = General educational diploma (high-school equivalent).

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2003, 2004, 2005 for households with incomes less than 185 percent of the Federal poverty line.

Table 1-9

Cross-classification of low-income households1 by food security status on ever-during-the-year scale and frequent/chronic scale

		Food security status based on frequent/chronic scale			
		Not frequently food insecure	Frequent or chronic low food security	Frequent or chronic very low food security	Total
		Percentage of all households			
Food security status based on ever-during-the-year scale	Food secure	75.11	0.00	0.00	75.11
	Low food security	16.25	1.73	0.00	17.98
	Very low food security	3.13	2.49	1.28	6.91
	Total	94.50	4.22	1.28	100.00

[1]Analysis sample included all households with incomes less than 185 percent of the poverty line and valid responses to all food security questions.

Characteristics of households in shaded cells are contrasted in logistic regression analysis presented in table 1-10.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2003, 2004, 2005 for households with incomes less than 185 percent of the Federal poverty line.

the year experienced frequent or chronic food insecurity (either low or very low). But a substantial proportion of households, represented by the shaded cells in table 1-9, constituted exceptions to these typical patterns. Households in these two cells either experienced only a short spell of very low food security, but were otherwise food secure (the lower-left shaded cell), or they experienced frequent or chronic low food security, but never experienced very low food security (the upper right shaded cell). All households in these atypical cells had raw scores of 6-10 on the trichotomous scale, and this raw score was controlled in the logistic regression analysis. I expected that households with more stable economic and demographic characteristics would predominate in the frequent-or-chronic-but-not-severe cell.

Differences in the characteristics of households in the two atypical cells were generally consistent with that expectation. Positive coefficients in table 1-10 indicate the log odds that households with the respective characteristic experienced severe but not frequent food insecurity. Negative coefficients indicate frequent or chronic food insecurity at a less severe level. The rightmost column indicates the observed percentage of the analysis sample in the severe-but-not-frequent category. Comparison to the national average of 64.5 percent provides an intuitively accessible indication of the bivariate association of that characteristic with the tradeoff between severity and frequency. When interpreting these associations it is important to keep in mind that the analysis sample comprises only households in these two atypical cells.

Food insecurity was generally more likely to be frequent or chronic but not severe for lower income households, while households with higher income were more likely to experience shorter, more severe episodes. Higher income is generally protective against food insecurity, so when food insecurity does occur, it usually results from a rapid or unexpected change in income or needs (Nord and Brent, 2002).

Reliance on retirement income and presence of elderly were each independently associated with much lower likelihood of short-but-severe episodes of food insecurity, consistent with the greater stability of retirement income.

Table 1-10

Logistic regressions of severe-but-not-frequent food insecurity on household economic and demographic characteristics among low-income households with either severe-but-not-frequent food insecurity or frequent-but-not-severe food insecurity[1]

Characteristic	Logistic regression model		Percentage of households with severe, but not frequent, food insecurity (not regression adjusted)
	Coefficient	p	Percent
All households in the analysis sample			64.5
Intercept	1.99	<.001	
Household income:			
Less than 50 percent of poverty line	-.50	.025	61.8
50-75 percent of poverty line	-.43	.056	63.0
75-100 percent of poverty line	-.41	.061	62.0
100-150 percent of poverty line	-.37	.078	65.2
150-185 percent of poverty line (reference)			74.4
Household employment:			
One or more adults employed full time (reference)			69.9
All adults retired, not in labor force	-.48	.073	41.9
Part-time worker for non-economic reasons, no full-time worker	-.07	.758	67.9
Part-time worker wants to work full time, no full-time worker	.08	.785	71.7
Unemployed, no employed adult	.33	.130	75.9
Disabled, no adult in labor force	-.60	<.001	54.4
Other, no adult in labor force	-.13	.605	63.9
Highest education level of adult			
Less than high school	-.59	<.001	51.3
High school or GED (reference)			65.4
Some college, no 4-year degree	.12	.409	70.5
Bachelor, other 4-year degree	.46	.096	77.7
Graduate or professional degree	-.09	.835	NA
Household structure:			
Two parents with child (reference)			68.5
Female with child, no spouse	-.26	.166	62.5
Male with child, no spouse	-.27	.368	59.2
Other household with child	.26	.731	NA
Two or more adults, no child	.02	.920	62.0
Woman living alone	.33	.139	59.3
Man living alone	.70	.002	74.0
Any elderly person in household	-.84	<.001	40.0
No elderly person in household (reference)			67.9
Race/ethnicity			
White non-Hispanic (reference)			64.0
Black non-Hispanic	.29	.055	66.3
Hispanic	.21	.269	65.1
Other	-.29	.259	57.7

—Continued

Table 1-10

Logistic regressions of severe-but-not-frequent food insecurity on household economic and demographic characteristics among low-income households with either severe-but-not-frequent food insecurity or frequent-but-not-severe food insecurity[1]—Continued

	Logistic regression model		Percentage of households with severe, but not frequent, food insecurity (not regression adjusted)
Characteristic	Coefficient	p	Percent
Citizen household reference person (reference)			64.4
Non-citizen household reference person	-.26	.269	65.0
Homeowner	-0.26	0.051	59.1
Renter or other tenure (reference)			66.7
Moved since entering Current Population Survey	.23	.251	69.4
Did not move since entering Current Population Survey (reference)			63.9
Metropolitan area residence			
In principal city	-.17	.259	67.9
Metropolitan, suburban or exurban (reference)			69.1
Metropolitan, principal city residence not identified	-.47	.007	61.2
Outside metropolitan area	-.57	<.001	55.1
Census Region			
Northeast (reference)			66.6
Midwest	-.02	.936	65.9
South	-.06	.748	62.9
West	-.13	.506	64.6
Raw score on polytomous scale			
Raw score = 6 (reference)			72.7
Raw score = 7	-.01	.930	73.1
Raw score = 8	-.64	<.001	60.6
Raw score = 9	-.60	.002	58.0
Raw score = 10	-1.73	<.001	34.9
Somers' D	.405		
Number of cases	1,668		1,668

[1]The analysis sample included households with incomes less than 185 percent of the poverty line and valid responses to all food security questions and one of two contrasting categories based on cross classification of the ever-during-the-year scale and the frequent/chronic scale: Either:

- Very low food security on the ever-during-the-year scale and food secure on the frequent/chronic scale, or
- Low (but not very low) food security on both scales (see table 1-9).

The coefficients indicate the log-odds of being in the former category; thus higher values indicate higher probability of severe but not frequent food insecurity and lower values indicate higher probability of chronic or frequent, but less severe, food insecurity.

NA = Not reported, 10 or fewer cases in one or more category.

GED = General educational development (high-school-equivalent diploma).

Shaded characteristics are significantly associated (90-percent confidence) with severe-but-not-frequent food insecurity with controls for other characteristics and for raw score on the polytomous scale comprising the sum of raw scores of the ever-during-the-year and frequent/chronic scales.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2003, 2004, 2005 for households with incomes less than 185 percent of the Federal poverty line.

The two categories overlap heavily, so the unadjusted percentages for both categories reflect the combined associations of retirement and elderly. Those percentages of households with short-but-severe episodes of food insecurity are by far the lowest of any measured category.

Food insecurity in households with no adult in the labor force and at least one adult out of the labor force because of disability was also more likely to be frequent or chronic but not severe. The earlier analysis indicated that this condition was positively and strongly associated with food insecurity and with chronic food insecurity (see table 1-7). The present analysis completes the picture, indicating that although households with disability were more likely to be food insecure, their experience of food insecurity was more likely to be temporally stable than that of otherwise similar households with no disability and with the same raw score on the trichotomous scale.

Associations with educational attainment also follow the expected pattern. Food insecurity was more likely to be frequent or chronic rather than severe in households with no high school educated adult. Food insecurity was less likely to occur at all for households with a college graduate (see table 1-7), but more likely to be severe and of short duration if the combination of severity and frequency places it in the analysis sample for table 1-10.

Homeowners were less likely to be food insecure than renters (see table 1-7), and if they were food insecure in one of the two atypical cells, they were more likely to experience frequent or chronic, but less severe food insecurity, consistent with their more stable economic condition (see table 1-10).

Households in suburban and exurban areas are more likely to experience short severe episodes of food insecurity, while food insecurity is more likely to be less severe but frequent or chronic for households in nonmetropolitan areas. Like higher income households, those in suburban and exurban areas are generally better off and more likely to experience food insecurity under exceptional circumstances and, therefore, as a single short episode.

Conclusions

The food security response data fit a polytomous Rasch model reasonably well, based on item-fit statistics. At the same time, the data also fit dichotomous Rasch models reasonably well when coded either as “ever versus never” or as “frequent/chronic versus never/occasional.” Based on item-fit alone, either approach is acceptable, and neither is clearly preferred.

However, findings in this study provide little support for adoption of a polytomous model for routine food security monitoring. The polytomous model would, in theory, provide more precise measurement, but at some expense in loss of transparency and ease of explanation. However, gains in precision (assessed by reduction in measurement error) at the levels of severity monitored by USDA would be small—perhaps near zero. Furthermore, there is some evidence that the greatest severity experienced during a year and the greatest severity experienced frequently or chronically during the year represent distinct dimensions in the response data. Responses in the highest frequency-of-occurrence category are positively correlated across items, conditional on the raw score on the polytomous model, and food insecurity ever-during-the-year and frequently-or-chronically-during-the-year not as highly correlated as would be expected if item responses at “ever” and “frequent” thresholds were conditionally independent. These findings call into question the statistical appropriateness of the polytomous model and the ordinality of raw score from the polytomous model as a measure of a single latent trait.

Measures of ever-during-the-year and frequent/chronic food insecurity are differentially associated with a number of household characteristics, and the differences are substantial for some characteristics. Associations of household characteristics with food insecurity experienced as short-but-severe spells versus frequent/chronic-but-not-severe are generally consistent with theory, suggesting that the lack of correlations between the measures is not random, but rather indicates distinctly different experiences of food insecurity with different causal factors.

These findings suggest that little would be gained by adopting a polytomous model of the type assessed here for routine food security monitoring. A dichotomous scale has several important advantages in its transparency and ease of explanation. Furthermore, although the CPS-FSS collects frequency-of-occurrence information on all but one of the items in the food security scale, no other survey currently does so, and to do so would increase the respondent burden unacceptably in many surveys.

If USDA considers it important to provide more adequate information on frequency and duration of food insecurity, this might better be accomplished by adding the frequent/chronic scale to its measurement system. This would provide a measure of frequent or chronic food insecurity and of frequent or chronic very low food security. In conjunction with that change, consideration may be given to removing the three frequency-of-occurrence items from the current standard scale. One or more of those items might be replaced with an additional ever-during-the-year item of similar severity. The need for a replacement is particularly important for the item indicating that adults cut the size of meals or skipped meals in 3 or more months, since that item plays a key role in identifying households with very low food security.

Chapter 2.

Modeling Conditional Dependence of Frequency-of-Occurrence Items

Abstract

An assumption of the Rasch measurement model is that items are conditionally independent. Three pairs of items in the U.S. Household Food Security Scale violate this assumption. Each pair consists of a base item indicating whether a condition ever occurred during the reference period and a followup item indicating how frequently the condition occurred. In the original assessment of items and development of the food security scale, these items were treated as if each set constituted two independent dichotomies. The correct way to model such a pair of dependent items is as a trichotomy rather than as a pair of dichotomies. In this chapter, the practical implications of this violation are explored, and the food security scale items are assessed treating these sets of dependent items appropriately as trichotomies. Modeled as trichotomies, the item sets fit the Rasch measurement model well enough that they do not substantially distort measurement in the adult scale and child scale. However, in the household food security scale that combines adult and child items, the trichotomous items do not fit well and introduce slight distortions into the measure.

Background

The Committee on National Statistics (CNSTAT) in its review of USDA's food security measurement methods recommended that USDA consider an alternative measurement model, "Treating items with frequency followup questions appropriately, for example, as a single ordered polytomous item rather than as two independent questions." (NRC, 2006, p. 10).

The dependent items in question are three pairs of items in the food security scale, each consisting of an initial question (hereafter "base question") and a followup question. The base question asks whether a behavior ever occurred during the last 12 months:

- In the last 12 months, did you or other adults in your household ever cut size of your meals or skip meals because there wasn't enough money for food?
- In the last 12 months, did you or other adults in your household ever not eat for a whole day because there wasn't enough money for food?
- In the last 12 months, did any of the children ever skip a meal because there wasn't enough money for food?

The followup, which is administered only if the response to the base question is affirmative, asks "How often did that happen, almost every month, some months but not every month, or in only 1 or 2 months?" In calculating the scale, responses to the base question and followup are treated as two dichotomous items. The first item is coded 1 for a response of "yes" to the initial (or base) question and coded 0 for "no." The second item is coded 1 for a response of "almost every month," or "some months but not every month," to the followup question, and is coded 0 for a response of “in only 1 or 2 months” or for a response of “no” to the base question.

These items are mutually dependent. If the items are modeled as separate dichotomies, their dependency violates the Rasch-model assumption of conditional independence. Conditional independence means that responses to items are uncorrelated across households with the same level of the food insecurity latent variable. The CNSTAT recommendation is that USDA consider the alternative of modeling each such pair of items as a trichotomy, that is, as a single item with three levels. Modeling the items in this way would take appropriate account of their mutual dependence.[1]

USDA analysts have been aware of this statistical issue since the beginning of the food security measurement project. Hamilton et al. (1997b, p. 18) noted that the dependencies artificially depressed (improved) item-fit statistics for the dependent items, and added, “We examined several alternative models with these items modeled as trichotomies rather than the multiple dichotomies, but the basic results of the models did not change.” Nord and Fogarty (2000) also found that the practical effects of these dependencies on item scores were negligible and suggested a two-step method for fitting models to food security data to avoid biasing item fit statistics.

> *Research reported in this paper, as well as several studies not reported here, have uniformly found that item calibrations for food security items are affected only negligibly by this particular form of item codependence, but that fit statistics for the codependent items are biased downward substantially. Item calibrations are not substantially affected because frequent occurrence is so much more severe than “ever” occurrence that the censoring resulting from item dependence is negligible. Fit statistics are biased downward (i.e., toward less misfit) because extremely unlikely outliers for the two items are prevented by the dependency. Fit statistics for item sets with frequency follow-up items should be assessed by first excluding the frequency follow-ups, and then, in a separate run, including the frequency follow-ups and excluding the base questions. Unbiased item severities can also be estimated using this method (with metrics equated using the remaining common items), but this is not generally necessary. The evidence is quite strong that item severities will not be affected by these dependencies (Nord and Fogarty, 2000, p. 14).*

While the Nord-Fogarty method improves item assessment considerably, the model against which it assesses the fit of dependent pairs of items does not definitively assure that raw score is asymptotically ordinal with respect to the latent trait.

In this chapter, the extent to which the current standard food security measures are distorted by this specific statistical violation is explored. First, the extent to which household food security scale scores and food security status classifications are affected was examined. Then item-fit statistics modeling each pair of items as a trichotomy were assessed.[2]

[1]The methodological extension of the Rasch model that is appropriate for modeling polytomous (multi-level) food-security items is the “partial credit” model, because it was developed in educational testing for the case in which a test answer may be wrong, partially correct, or fully correct.

[2]This chapter focuses narrowly on the specific issue of the sets of dependent items currently included in the standard food security measures. The CNSTAT panel also suggested that USDA examine a model including polytomous items representing the full range of frequency-of-occurrence information that is available for all items in the food security scale. That assessment is the topic of Chapter 1 and is not explored further here.

Overview of Issues and Results

The CNSTAT panel is clearly correct in pointing out that these pairs of dependent items should be modeled as trichotomies. The principal issue of concern for this analysis is the assessment of how well the three sets of dependent items, and the remaining dichotomous items, fit the Rasch model when the dependent pairs of items are appropriately modeled as trichotomies. If USDA continues to classify household food security status as a discrete categorical condition based on raw score, then classification and prevalence estimates—the primary and most visible uses of the food security measures—will be identical whether each item pair is treated as two dichotomies or as a single trichotomy.

A secondary question is whether past estimates of household severity measures and estimated measurement errors were distorted in any important way because these dependencies were ignored. Findings in this study will show that the continuous household measures of food insecurity based on the two methods differ so slightly that the practical effects of having ignored the dependencies in the past were negligible for those measures. Measurement error was underestimated across most of the range of severity when dependencies were ignored (by up to 20 percent for some raw score groups). However, relatively little research use has been made of error estimates, and the true measurement error is essentially the same for both measures since they place exactly the same households in each raw score group and differ negligibly in their calculation of the mean severity within each raw score group.

The substantive question, then, is whether the items fit the polytomous Rasch model sufficiently well to justify modeling two thresholds for these items. This is the question to which almost all of this chapter is devoted. If the answer is "yes," then there are no implications for the standard measurement methods except to slightly revise the linear measure scores next time official guidance is published and to use the corrected estimates for measurement error for estimating measurement reliability and classification reliability. If the answer is "no," then alternatives to the current measures may be considered, possibly dropping the second item from one or more of the dependent item pairs.

Data and Methods

Primary analyses used data from the 2003, 2004, and 2005 Current Population Survey Food Security Supplements (CPS-FSS), the data source USDA uses for its annual food security monitoring report. Data for 3 years were combined to provide very stable estimates of item parameters and response patterns. Selected analyses were repeated using data from the 1995 and 1998 CPS-FSS data. The 1995 data were the basis for the initial development of the food security scale, and this reanalysis examines whether the conclusions of the original analysis would have differed if the dependent item pairs had been modeled appropriately as trichotomies. Furthermore, the 1995 data did not include internal screeners within the core module (which will be discussed later in this chapter), and thus avoids some of the analytic problems those screeners cause in the 1998 and later data. The 1998 data are reanalyzed because those data were the basis for the current guidance USDA provides to researchers who wish to implement the food security measure in their surveys (Bickel et al., 2000; Nord and Bickel, 2002).[3]

Rasch models treating the dependent sets of items as trichotomies were estimated using SAS programs developed by ERS to implement conditional maximum likelihood (CML) estimation methods based on the single-parameter partial credit Rasch model (Masters, 1982).[4] (Hereafter, these models are referred to as "trichotomous" models even though only three of the items were, in fact, trichotomous.) In general, CML estimation is superior to joint maximum likelihood (JML) estimation for small item sets such as those used to measure food security. JML-estimated item parameters are inconsistent—biased toward greater dispersion (Fisher and Molenaar, 1995). This bias is greater for scales consisting of small numbers of items, and distorts item-fit statistics in predictable ways (Nord, 2006).

Household measures for each of the 3 standard U.S. food security scales (18-item household scale, 10-item adult scale, and 8-item child scale) based on the trichotomous item parameters estimated from the 1998 CPS-FSS data were compared for linearity with the measures provided by Bickel et al. (2000) and Nord and Bickel (2002). Measurement error was compared between the trichotomous measures and naïve dichotomous measures. Household measures corresponding to each raw score were derived by maximum likelihood estimation from the estimated item parameters. Classifications are identical in the trichotomous and the naïve dichotomous models, since raw scores in the two models are identical.

The fit of all items (trichotomous as well as dichotomous) in the trichotomous models was assessed based in item-infit and item-outfit statistics. These are chi-square-type item misfit statistics that compare squared errors of item responses with squared errors expected under model assumptions. Infit, which is information weighted, is the more useful statistic and is sensitive to overall fit of the item (see box, Calculation of Item Fit Statistics). Outfit is sensitive to highly improbable (i.e., outlier) responses.

The primary assessments were based on the 2003-05 CPS-FSS data using the household supplement weights. Then several alternative analyses were conducted to confirm the robustness of the main results. The main analyses as well as each alternative were conducted separately for the 18-item household scale, the 10-item adult scale, and the 8-item child scale.[5] The alternative analyses included the following:

- Restricting the sample to households with incomes less than 185 percent of the Federal poverty line. This avoids any distortions due to the screener prior to the first of the food security questions, since households with incomes in this low-income range were not subject to that screener.
- Unweighted estimation. The main analyses used sampling weights (household supplement weights). In principle, weights do not affect Rasch model fit or parameter estimates, provided Rasch assumptions are met in the population. These analyses were conducted to verify that assumption. In fact, the unweighted estimates were so similar to the weighted estimates, that I do not present the unweighted estimates in this chapter.

[3]The current standard food security scales (household, adult, and child) are based on item parameters estimated from the 1998 CPS-FSS data. Each year, ERS assesses overall model fit, estimated item-severity parameters, and item-fit statistics in the CPS-FSS data to confirm that use of the standard method is still appropriate. Through 2006, the response patterns have remained sufficiently stable that it has not been deemed necessary to revise the 1998 household scale scores and classification specifications. An exception was a split-ballot test in the eighth rotation of the 2000 CPS-FSS, which substituted three nonstandard questions for the initial three questions in the scale. The response characteristics of the test items differed from those of the three items they replaced, and the public-use data for households in that test rotation reflect household food security scores and classifications based on the item scores estimated from those data.

[4]The partial-credit Rasch model allows the interthreshold distance (in logits) to differ between items. The SAS programs estimate item parameters in a data step using Newton-Raphson iterative approaches, following response-probability formulations described by Fischer and Molenaar (1995). Item-fit statistics are then calculated based on the final modeled probabilities for each cell in the raw-score-by-item matrix. The ERS programs have been tested against the SAS Logistic procedure with the new "Strata" command in multiple data sets using dichotomous items only and found to give identical results. They were also tested using simulated data with trichotomous items that were generated to be perfectly consistent with polytomous Rasch model assumptions; they recovered the generating parameters exactly and gave perfect item-fit statistics.

[5]The larger food security measurement assessment project (of which this report is a part) will explore, among other things, whether the 18-item household food security scale that includes both adult and child items should be abandoned in favor of the separate adult and child scales. The main analyses were conducted on all three scales in case the project eventuates in such a recommendation.

Calculation of Item-Fit Statistics

Item-infit is calculated as follows:

$$INFIT_i = SUM\,[(X_{i,h} - P_{i,h})^2] \,/\, SUM[P_{i,h} - P_{i,h}{}^2]$$

where:

$X_{i,h}$ is the observed response of household h to item i (1 if response is yes, 0 if response is no);

$P_{i,h}$ is the probability of an affirmative response by household h to item i under model assumptions, given the item calibration and the raw score of the household.

The expected value of the infit statistic for each item is 1.0 if the data conform to Rasch model assumptions. Values above 1.0 indicate that the item discriminates less sharply than the average of all items in the scale.

Item outfit is calculated as the average across households of the squared error divided by the expected squared error:

$$OUTFIT_i = SUM\,[(X_{i,h} - P_{i,h})^2 \,/\, (P_{i,h} - P_{i,h}{}^2)] \,/\, N$$

where:

$X_{i,h}$ is the observed response of household h to item i (1 if response is yes, 0 if response is no);

$P_{i,h}$ is the probability of an affirmative response by household h to item i under Rasch assumptions, given the item calibration and the raw score of the household;

N is the number of households.

The expected value of each item's outfit statistic is 1.0 if the data conform to Rasch model assumptions. Values above 1.0 indicate a higher than expected proportion of "erratic" responses—affirmative responses to a severe item by households that affirmed few other items or denials of a low-severity item by households that affirmed many other items.

- JML estimation. The main analyses were repeated using JML estimation methods implemented in Winsteps. Results did not differ in any important way from the CML results (taking into consideration known distortions in JML estimates), and the JML results are not presented in this chapter.

- Accounting for, or removing, the effects of screening within the food security core module. To reduce respondent burden, the food security questions are administered in three blocks, with the items ordered approximately by severity—the least severe items presented first. If there are no affirmative responses to any question in block 1, then blocks 2 and 3 are skipped, and negative responses are imputed. Similarly, if there are no affirmative responses to any of the questions in block 2, then block 3 is skipped, and negative responses are imputed. These screens artificially improve the fit of the response data to the Rasch model, since some low-probability responses cannot appear in the data. In effect, households that would have given such improbable responses had other, more probable, responses imputed. For the household and adult scales, modified CML estimation methods were used that take account of the internal screens. This is accomplished by omitting response patterns that would be obviated by the screens when calculating expected response proportions in each raw score. This method could not be used for the child scale because of specific characteristics of the screening. Instead, the estimation sample was restricted to households not affected by screening; households were included only if they would have been screened into the second and third blocks based on responses to adult items in the earlier blocks.

- Collapsing each trichotomy to a single ever/never dichotomous item. Response to the "how often" followup questions was ignored, and the items were coded based on the response to the yes/no base question. This examined the extent to which the fit of all items might be improved by omitting the "how often" followup items—a possible alternative to the current standard methods if the trichotomous items are found to not fit the Rasch model acceptably well.

Households with no valid food security responses were excluded from the analysis (0.3 percent of households) as were households with missing responses to any of the base questions or frequency followup questions on which the scale is based (0.6 percent of all households and 2.3 percent of households with nonextreme responses). Extreme responses (no affirmative responses or all affirmative responses) were omitted from the analyses since these do not contribute information about the relative severity of items or the item response characteristics. Households with no affirmative responses comprised over 80 percent of the sample, and households with all affirmative responses comprised about a half of 1 percent. The 2003-05 analysis samples after these exclusions consisted of:

- 24,902 households (with or without children) that were nonextreme on the 8-item adult scale (which includes 2 trichotomies and thus has a maximum raw score of 10)
- 7,814 households with children that were nonextreme on the 7-item child scale (which includes one trichotomy and thus has a maximum raw score of 8)
- 12,581 households with children that were nonextreme on the 15-item household scale (which includes 3 trichotomies and thus has a maximum raw score of 18)
- 25,694 households (with or without children) that were nonextreme on the 15-item household scale if children were present or nonextreme on the 8-item adult scale if no children were present

The 1995 and 1998 samples were about one-third the size of the 3-year 2003-05 samples.

An important analytic issue is the interpretation of item-fit statistics. How good is good enough? There is no hard-and-fast rule, but the following guidance has been offered in various publications:

Wright and Lineacre (1994) provide two tables describing “reasonable mean-square fit values” (tables 2-1 and 2-2). The tables are not quite identical, but taken together suggest that for survey-based measurement, item-fit values in the range of 0.6 to 1.4 may be acceptable. Items with infits somewhat higher than 1.4—perhaps as high as 2.0—may not substantially degrade measurement performance, but will not contribute positively to measurement.

Bond and Fox (2001) reproduce the first table from Wright and Lineacre.

Hamilton et al. (1997b) in their initial work developing the food security measure applied a rather stringent standard: “Generally speaking, a mean square fit statistic that is greater than 1.20 indicates a poor fitting item, whereas a mean square fit statistic that is less than .80 indicates an item that is redundant with other similar types of items in the scale.” This would be consistent with the standard that Wright and Lineacre suggest for high-stakes multiple-choice question tests.

For purposes of the present assessment, I propose a reasonably stringent standard, although less stringent than that suggested by Hamilton et al. Considering that the food security measure plays a highly visible role as an indicator of material well-being, it should be held to a somewhat higher standard than routine survey-based measurement. I suggest that for large-sample multiple-year assessment of items, we should expect item-infit statistics in the range of 0.7 to 1.3.[6] For assessment of usability of data in a single survey, the standard may be relaxed somewhat to 0.6 to 1.4. Previous analysis using simulated data suggest that differences in item discrimination within this range bias prevalence estimates by relatively small proportions (Nord, 2006).

Primary attention in these assessments is given to the item-infit statistic. While item-outfit statistics are also examined, they have several limitations that should be kept in mind when interpreting them. In practice, outfit statistics are very sensitive to a few highly unexpected observations. As few as two

Table 2-1
Reasonable item mean-square ranges for “infit” and “outfit”

Type of test	Range
MCQ (high stakes) [multiple-choice question test]	0.8 – 1.2
MCQ (run of the mill) [multiple-choice question test]	0.7 – 1.3
Rating scale (survey)	0.6 – 1.4
Clinical observation	0.5 – 1.7
Judged (agreement encouraged)	0.4 – 1.2

Source: Wright and Linacre, 1994.

Table 2-2
Interpretation of parameter-level mean-square fit statistics

> 2.0	Distorts or degrades the measurement system
1.5 – 2.0	Unproductive for construction of measurement, but not degrading
0.5 – 1.5	Productive for measurement
< 0.5	Less productive for measurement, but not degrading. May produce misleadingly good reliabilities and separations

Source: Wright and Linacre, 1994.

[6]In the past, I have used (and recommended to other researchers) a somewhat more stringent standard as follows: “Item infits in the range of 0.8 to 1.2 are considered good and that infits up to 1.3 are acceptable but indicate items that should be improved prior to widespread use. Infits below 0.8 indicate items that are more strongly associated with the condition measured by the entire set of items. Including such items is not problematic, but they are undervalued in the equal-weighted Rasch measure.” My current view, based on further reading of the literature and my own studies using simulated data, is that the 1.2 standard may be a bit too stringent. Reduced discrimination of a single item to a degree that produces an infit of 1.3 results in a relatively small bias in prevalence estimates calculated from raw-score-based classification in simulated data (Nord, 2006).

or three highly unexpected responses among several thousand households can elevate the outfit for that item to 10 or higher. Furthermore, outfit statistics are seriously distorted by the screening implemented in many surveys, including the CPS-FSS, beginning in 1998. To reduce respondent burden and annoyance, households that deny all less severe items are skipped over the remaining, more severe, items. This obviates certain improbable response patterns, which tends to suppress outfit statistics for the more severe items and inflate them for less severe items (the latter because overall model fit is artificially improved by the screening; Nord, 2006). Carefully interpreted, outfit statistics may help identify items that present cognitive problems or have idiosyncratic meanings for small subpopulations, but I am reluctant to apply specific cutoffs for assessment.

Finally, the effect of the misfit of two of the trichotomous items on prevalence rates of overall food insecurity and of very low food security is estimated. Two of the trichotomous items are found to have unacceptably high fit statistics by conventional criteria. The effects of these item misfits on prevalence rates were estimated by comparing observed with expected responses in level 2 (occurrence in more than 1 or 2 months) of each item in "rest of the items" raw score 2, just below the food insecure threshold, and raw score 7, just below the threshold for very low food security. Raw scores for this purpose were calculated with level 2 of the item under assessment omitted from the calculation. This analysis does not assess the full effects of the misfit of these items, but focuses on the effects with the greatest practical importance in the highly visible food security monitoring process.

Typical item-severity parameters for the household, adult, and child scales are provided in appendix B along with the food security status classification specifications for each scale.

Findings

Household Measures

The rank order of households' measured food insecurity is the same whether it is based on the naïve dichotomous model (which ignores the dependence between base and followup items) or on the trichotomous model. In either case, a response of "no" contributes 0 points to the raw score; a response of "yes, but in only 1 or 2 months" contributes 1 point; and a response of "yes, in some months but not every month" or "yes, in almost every month" contributes 2 points to the raw score. Since the standard method bases household food security status classification on raw score, those classifications would be the same for either the naïve dichotomous model or the technically correct trichotomous model. Clearly, then, provided that the same raw-score thresholds are used, household food security status classification under the trichotomous model is identical to that based on the current standard methods. It is important to note, however, that although the two methods yield the same classification, the ordinality of raw score with respect to the latent trait of food insecurity, upon which the validity of that classification rests, is only assured if the data fit the trichotomous Rasch model.

In principle, modeling the dependent pairs of items as trichotomies provides correct household measures of the latent trait. In practice, however, the effect of ignoring these dependencies in the case of the food security scale is negligible. Household measure values in the USDA *Guide to Measuring Household Food Security, Revised 2000* (Bickel et al., 2000) differ only very slightly from measures based on the same 1998 data with the three pairs of dependent items modeled as trichotomies (fig. 2-1). Corresponding analyses of the 10-item adult scale and 8-item child scale (not shown) indicated even smaller distortions for those scales.

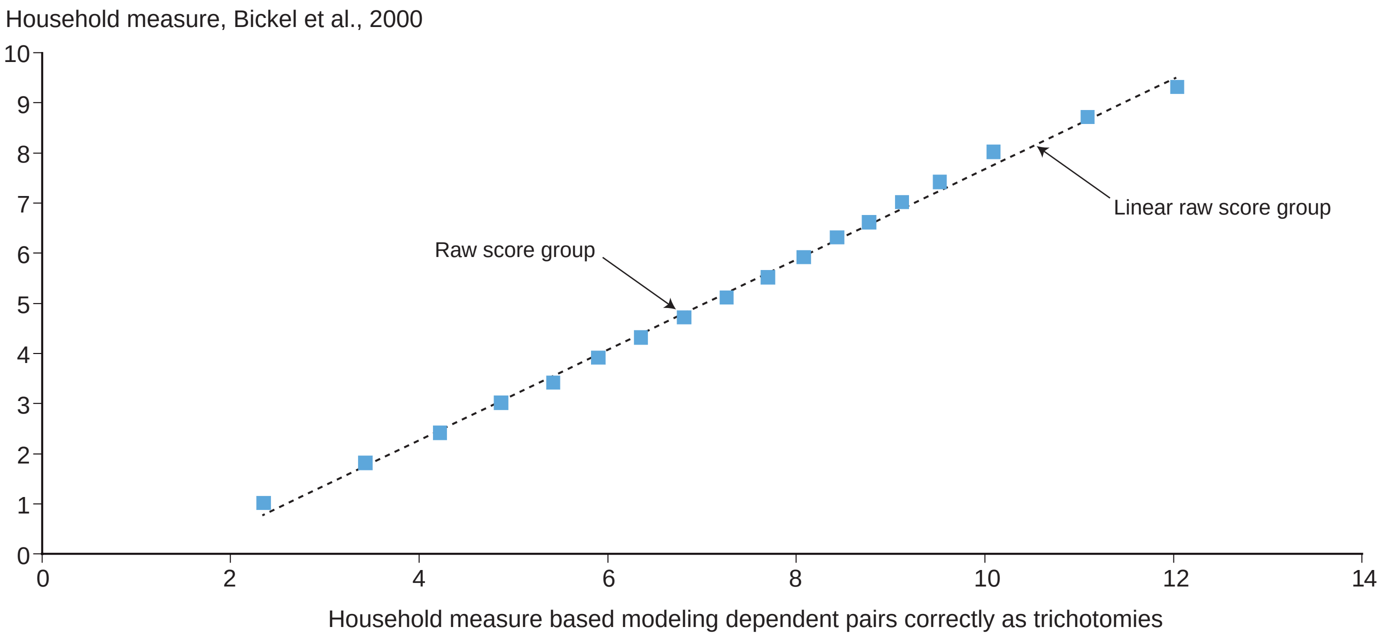
Ignoring the dependencies between these pairs of items does, however, result in underestimates of measurement error (the standard deviation of "true" severity of food insecurity across households within each raw score) across much of the measured range for all three scales. Underestimates of measurement error averaged about 8 percent for the household scale, and were as high as 20 percent for some score groups; average underestimates were 10 percent for the adult scale and 3 percent for the child scale. These error estimates are not widely used by researchers, but they play a role in estimating classification reliability and estimating the extent of bias due to differential item function between population subgroups. Improving the accuracy of the error estimates for these purposes is a further reason to model these item-pairs properly as trichotomies.

Item-Fit Assessment

The central question for this chapter is: Do the dependent pairs of items, when correctly modeled as trichotomies, fit the Rasch model well enough to justify basing household measures and food security classification on their raw scores?

Figure 2-1

Comparison of household measures for households with children, USDA standard (Bickel et al., 2000) versus model that treats the three dependent pairs of items as trichotomies



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplement of August 1998 for households with children.

Household scale (combined adult and child items)—Item-infit statistics for the 15-item household scale are presented in table 2-3 and item-outfit statistics in table 2-4. The two adult-referenced trichotomous items *(Adult cut size of meal or skipped meal and Adult did not eat for whole day)* have high infit statistics in most of the models—above acceptable limits in many of the models. The child-referenced trichotomy *(Child skipped meal)*, on the other hand, had quite good infit values in all models and may be considered unproblematic in the 15-item combined adult-child measure.

The infit statistic for the item *Adult cut size of meal or skipped meal* was unacceptably high in the 2003-2005 data (model 1) and barely within the acceptable range in the 1998 data (model 5). When households without children were included in the estimation, with the child-items coded as missing, the infit of *Adult cut size of meal or skipped meal* was somewhat better (model 2). This reflects the more consistent relationship of the item to the other adult-referenced items than to the child items. About two-thirds of households have no children, so the relationships with adult-referenced items predominate in calculating the fit statistics in this model. Restricting the

Table 2-3

Item-infit statistics, 15-item household scale

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Item	Households with child 2003-05	All households 2003-05	Households with child and income < 185-percent-poverty 2003-05	Households with child 1995	Households with child 1998	Households with child 2003-05 adjust for effect of screens[1]	Households with child 2003-05 trichotomous items as dichotomies[2]
Worried food would run out	1.04	1.06	1.05	1.07	1.04	1.03	1.04
Food bought did not last	.99	.97	.98	.92	1.01	.97	.99
Could not afford to eat balanced meals	1.07	1.18	1.06	.91	1.04	1.05	1.08
Few kinds of low-cost food for children	1.08	1.03	1.09	1.12	1.10	1.05	1.08
Could not feed children balanced meals	.91	.89	.92	.96	.93	.90	.92
Children not eating enough	1.02	1.01	1.05	.96	1.01	1.01	1.07
Adult cut size of meal or skipped meal[3]	1.35	1.28	1.32	1.52	1.30	1.33	1.02
Respondent ate less than felt should	.86	.79	.84	.92	.89	.87	.94
Respondent hungry but did not eat	.88	.85	.86	.88	.89	.87	.97
Respondent lost weight	.89	.90	.91	.96	.97	.88	.99
Adult did not eat for whole day[3]	1.48	1.33	1.43	1.57	1.43	1.47	1.05
Cut size of child's meal	.94	.93	.93	.89	.89	.98	1.02
Child skipped meal[3]	1.13	1.14	1.19	1.02	1.06	1.13	.89
Child hungry, could not afford more food	.83	.83	.84	.84	.75	.85	.84
Child did not eat whole day	.95	.95	.94	.97	.95	.96	.97
Estimation sample size	12,581	25,694	7,874	4,285	3,905	12,577	12,578

[1]The conditional maximum likelihood (CML) estimation program used in model 6 omits response patterns that would be obviated by the screens implemented at data collection when calculating expected response proportions in each raw score.

[2]In model 7, the followup questions ("How often did that happen?") were omitted. The three items coded as trichotomies in the other models were coded based only on yes/no response to base questions.

[3]Items were modeled as trichotomies (except in model 7).

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements.

Table 2-4

Item-outfit statistics, 15-item household scale

Item	Model 1 Households with child 2003-05	Model 2 All households 2003-05	Model 3 Households with child and income < 185-percent-poverty 2003-05	Model 4 All households 1995	Model 5 All households 1998	Model 6 All households 2003-05 adjust for effect of screens[1]	Model 7 All households 2003-05 trichotomous items as dichotomies[2]
Worried food would run out	1.59	1.84	1.38	1.24	1.73	1.46	2.11
Food bought did not last	1.07	1.13	1.02	.85	1.28	1.03	1.17
Could not afford to eat balanced meals	1.21	1.46	1.18	.85	1.15	1.15	1.29
Few kinds of low-cost food for children	1.54	1.42	1.91	1.15	1.21	1.43	1.96
Could not feed children balanced meals	.78	.75	.79	.77	.80	.77	.80
Children not eating enough	.64	.62	.70	.75	.70	.79	.72
Adult cut size of meal or skipped meal[3]	1.34	1.17	1.38	2.17	1.30	1.64	.87
Respondent ate less than felt should	.63	.59	.63	.90	.67	.79	.71
Respondent hungry but did not eat	.59	.56	.58	.66	.54	.73	.69
Respondent lost weight	.51	.51	.60	.69	.44	.58	.64
Adult did not eat for whole day[3]	.92	.78	1.13	2.21	.93	2.74	.53
Cut size of child's meal	.40	.39	.43	.76	.35	.94	.47
Child skipped meal[3]	.80	.77	.88	.88	.55	2.26	.39
Child hungry, could not afford more food	.29	.28	.32	.56	.23	.75	.32
Child did not eat whole day	.40	.39	.26	2.87	.40	.85	.59
Estimation sample size	12,581	25,694	7,874	4,285	3,905	12,577	12,578

[1]The conditional maximum likelihood (CML) estimation program used in model 6 omits response patterns that would be obviated by the screens implemented at data collection when calculating expected response proportions in each raw score.

[2]In model 7, the followup questions ("How often did that happen?") were omitted. The three items coded as trichotomies in the other models were coded based only on yes/no response to base questions.

[3]Items were modeled as trichotomies (except in model 7).

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements.

sample to low-income households (model 3) and accounting for the effects of the intra-module screens in the model-fitting process (model 6) improved the infit only slightly. The infit of *Adult cut size of meal or skipped meal* in the 1995 data (model 4) was substantially higher than in later years. The question order was somewhat different (less intuitively sensible, perhaps) in the 1995 survey, which may have been a factor.

Item-outfit statistics for this item follow a pattern very similar to the infit statistics (table 2-4). The distortion of outfit statistics by the intra-module screening is apparent in these models. With the exception of model 4 (the 1995 data, which had no such screening), and model 6 (in which screening was taken into account in the model fitting), outfits of the least severe items tend to be elevated, and those of the most severe items depressed.

It appears, then, that if the combined adult-child household scale is retained, inclusion of the second item (or the trichotomous coding) for *Adult cut size of meal or skipped meal* may not be justified. Omitting that threshold by replacing the trichotomous item with a single dichotomous item based on the yes/no base question gives a very acceptable infit of 1.02 and outfit of 0.87 (model 7). This suggests that the high infit of the trichotomous item was primarily due to poor discrimination at the upper threshold of the item (i.e., between occurrence in only 1 or 2 months and more frequently). This was confirmed by calculations (not shown) of separate infit statistics for the two thresholds in the trichotomous model.

The infit statistics for the item *Adult did not eat for whole day* followed a pattern similar to that of *Adult cut size of meal or skipped meal*, but were even more problematic. The outfit statistic for this item is strongly affected by screening, as it follows the second of the two internal screens. Thus, the outfits are near or below unity in models 1, 2, 3, and 5, but are relatively high in models 4 and 6 when screens are absent or their effects removed methodologically. This item, like *Adult cut size of meal or skipped meal*, performs quite well as a dichotomy (model 7) and should probably be limited to such if the combined adult-child scale is retained.

Item-infit statistics for the 12 dichotomous items were all well within an acceptable range in all of the models. Infit statistics for almost all of those items improved somewhat or were essentially unchanged when the three trichotomous items were collapsed to dichotomies (model 7). Outfit statistics are so distorted by screening that only models 4 and 6 provide directly interpretable information. Outfits of most of the dichotomous items in those models are quite good. The outfits of *Worried food would run out* and *Few kinds of low-cost food for children* were somewhat high in 2003-05, but not in 1995. The outfit of Child did not eat for whole day was quite high in 1995, but low in 2003-05. Analysis of the 1995 data (not shown) indicates that the high outfit for this item reflects just a single highly improbable response. One household reported this item, but denied all other food security scale items.

Adult scale (8 adult-referenced items)—Item-infit statistics for the 8-item adult scale are presented in table 2-5 and item-outfit statistics in table 2-6. The two trichotomous adult items fit Rasch-model assumptions considerably better in this scale than in the scale with both adult and child items. The similarity of the infit statistics for the two trichotomous items in the subsample of households with children (model 1) and the full sample (model 2) confirms that the differences between the corresponding models in the household scale (table 2-3) resulted from the effects of the child items, not from differences in response patterns to the adult items in the two samples.

Item-infit statistics for *Adult cut size of meal or skipped meal* were in the acceptable range in all models except the 1995 data (model 4). The higher infit in 1995 does not appear to have been a result of the lack of screening in that year, as the infit in model 6, which accounted for the effects of screening, was quite good. Outfit statistics for this item were also reasonably good except for the 1995 data (model 4), in which the outfit was somewhat high.

Item-infit statistics for *Adult did not eat for whole day* were within the acceptable range for all models. Outfit statistics are too distorted by screening to be meaningful except in models 4 and 6, in which they were somewhat high.

The six dichotomous items in the adult scale had infit statistics well within the acceptable range in all models, including model 7, in which the two

Table 2-5

Item-infit statistics, 8-item adult scale

Item	Model 1 Households with child 2003-05	Model 2 All households 2003-05	Model 3 Households income < 185-percent-poverty 2003-05	Model 4 All households 1995	Model 5 All households 1998	Model 6 All households 2003-05 adjust for effect of screens[1]	Model 7 All households 2003-05 trichotomous items as dichotomies[2]
Worried food would run out	1.00	1.01	1.00	1.08	1.03	0.99	1.01
Food bought did not last	.90	.92	.90	.84	.95	.89	.92
Could not afford to eat balanced meals	1.23	1.25	1.25	1.08	1.16	1.21	1.29
Adult cut size of meal or skipped meal[3]	1.16	1.17	1.19	1.47	1.13	1.16	.84
Respondent ate less than felt should	.74	.73	.73	.82	.77	.76	.78
Respondent hungry but did not eat	.83	.82	.81	.79	.80	.82	.89
Respondent lost weight	.85	.88	.88	.90	.82	.87	.97
Adult did not eat for whole day[3]	1.20	1.20	1.20	1.24	1.18	1.22	.98
Estimation sample size	11,789	24,902	14,689	7,577	7,330	24,768	24,756

[1]The conditional maximum likelihood (CML) estimation program used in model 6 omits response patterns that would be obviated by the screens implemented at data collection when calculating expected response proportions in each raw score.

[2]In model 7, the followup questions ("How often did that happen?") were omitted. The two items coded as trichotomies in the other models were coded based only on yes/no response to base questions.

[3]Items were modeled as trichotomies (except in model 7).

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements.

Table 2-6

Item-outfit statistics, 8-item adult scale

Item	Model 1 Households with child 2003-05	Model 2 All households 2003-05	Model 3 Households income < 185-percent-poverty 2003-05	Model 4 All households 1995	Model 5 All households 1998	Model 6 All households 2003-05 adjust for effect of screens[1]	Model 7 All households 2003-05 trichotomous items as dichotomies[2]
Worried food would run out	1.50	1.77	1.97	1.40	1.84	1.62	2.38
Food bought did not last	1.25	1.17	1.21	.81	1.30	1.12	1.47
Could not afford to eat balanced meals	1.66	1.66	1.72	1.30	1.68	1.57	2.27
Adult cut size of meal or skipped meal[3]	1.17	1.07	1.14	1.87	1.06	1.26	0.72
Respondent ate less than felt should	.57	.55	.57	.77	.60	.70	.61
Respondent hungry but did not eat	.50	.52	.52	.57	.50	.65	.61
Respondent lost weight	.52	.49	.55	.84	.36	.56	.61
Adult did not eat for whole day[3]	.73	.68	.73	1.41	.66	2.04	.50
Estimation sample size	11,789	24,902	14,689	7,577	7,330	24,768	24,756

[1]The conditional maximum likelihood (CML) estimation program used in model 6 omits response patterns that would be obviated by the screens implemented at data collection when calculating expected response proportions in each raw score.

[2]In model 7, the followup questions ("How often did that happen?") were omitted. The two items coded as trichotomies in the other models were coded based only on yes/no response to base questions.

[3]Items were modeled as trichotomies (except in model 7).

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements.

trichotomous items were collapsed to dichotomies. Outfit statistics were, again, seriously distorted by internal screening. In the two nonscreened models (models 4 and 6), outfits of the dichotomous items were reasonably good, although outfits were somewhat high for *Worried food would run out* in both models and for *Could not afford balanced meals* in model 6.

In the adult scale, then, unlike in the household scale, entering *Adult cut size of meal or skipped meal* and *Adult did not eat for whole day* as trichotomies appears to be justified by their fit statistics. The current standard method of scoring an additional raw score point for reported recurrence of these conditions in more than 1 or 2 months is consistent with the fit of the items.

Child scale (7 child-referenced items)—Item-infit statistics for the 7-item child scale are presented in table 2-7 and item-outfit statistics in table 2-8. Infit statistics for the only trichotomous child item, *Child skipped meal*, were within the acceptable range in all models as were those of the six dichotomous items.

Outfit statistics, even in the two nonscreened models (models 4 and 6), were quite high for some items. The child food security scale is somewhat unusual in that the items discriminate so strongly (responses are so consistently ordered in accordance with the severity of the items) that a very small number of improbable responses results in a high outfit. Evidence of the high discrimination of the child items include:

Table 2-7

Item-infit statistics, 7-item child scale

Item	Model 1 Households with child 2003-05	Model 2	Model 3 Households with child income < 185-percent-poverty 2003-05	Model 4 Households with child 1995	Model 5 Households with child 1998	Model 6 Households with child 2003-05 avoid effect of screens[1]	Model 7 Households with child 2003-05 trichotomous items as dichotomies[2]
Few kinds of low-cost food for children	1.03	Model 2 is not relevant for assessment of the child scale, but numbering of models was retained as in earlier tables for ease of comparability	1.03	1.05	1.03	1.03	1.03
Could not feed children balanced meals	.80		.77	.93	.82	.73	.80
Children not eating enough	.80		.84	.77	.90	.78	.80
Cut size of child's meal	.94		.94	.96	.92	.98	1.01
Child skipped meal[3]	1.17		1.22	1.18	1.15	1.15	.87
Child hungry, could not afford more food	.85		.84	.85	.79	.86	.89
Child did not eat whole day	.97		.96	1.02	.98	.97	1.00
Estimation sample size	7,814		5,196	2,557	2,393	3,365	7,812

[1]In order to obviate the effect of screening within the module, cases were included in model 6 only if responses were positive to at least one adult item in block 1 (*Worried food would run out, Food bought did not last,* and *Could not afford to eat balanced meals*) and at least one adult item in block 2 (*Adult cut size of meal or skipped meal, Respondent ate less than felt should, Respondent hungry but did not eat,* and *Respondent lost weight*). This ensures that all households in the estimation sample were asked all child questions.

[2]In model 7, the followup question ("How often did that happen?") was omitted. The item coded as a trichotomy in the other models was coded based only on yes/no response to base questions.

[3]Item was modeled as a trichotomy (except in model 7).

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements.

Table 2-8

Item-outfit statistics, 7-item child scale

Item	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	Households with child 2003-05	Model 2 is not relevant for assessment of the child scale, but numbering of models was retained as in earlier tables for ease of comparability	Households with child income < 185-percent-poverty 2003-05	Households with child 1995	Households with child 1998	Households with child 2003-05 avoid effect of screens[1]	Households with child 2003-05 trichotomous items as dichotomies[2]
Few kinds of low-cost food for children	6.23		7.70	1.79	2.49	5.76	8.01
Could not feed children balanced meals	1.02		.98	1.26	1.16	.94	1.11
Children not eating enough	1.01		1.07	.66	1.18	1.10	1.22
Cut size of child's meal	2.04		2.28	1.33	1.45	2.26	2.21
Child skipped meal[3]	4.85		5.56	3.42	2.59	5.36	2.73
Child hungry, could not afford more food	1.70		1.14	1.31	1.23	1.79	1.88
Child did not eat whole day	8.91		5.64	3.06	1.15	9.49	12.13
Estimation sample size	7,814		5,196	2,557	2,393	3,365	7,812

[1]In order to obviate the effect of screening within the module, cases were included in model 6 only if responses were positive to at least one adult item in block 1 (*Worried food would run out, Food bought did not last,* and *Could not afford to eat balanced meals*) and at least one adult item in block 2 (*Adult cut size of meal or skipped meal, Respondent ate less than felt should, Respondent hungry but did not eat,* and *Respondent lost weight*). This ensures that all households in the estimation sample were asked all child questions.

[2]In model 7, the followup question ("How often did that happen?") was omitted. The item coded as a trichotomy in the other models was coded based only on yes/no response to base questions.

[3]Item was modeled as a trichotomy (except in model 7).

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements.

- Responses of 82 percent of the households in model 7 with nonextreme responses were ordered exactly consistently with the severity of the items (i.e., were perfectly consistent with a Guttmann model).
- The average discrimination of the child-referenced items, as measured by standard deviation of the item calibrations, was 23 percent higher when the items were modeled alone than when they were modeled jointly with the adult items (analysis not shown).
- Outfit statistics for the child items were much lower when modeled jointly with the adult items (compare model 6 in tables 2-4 and 2-8).

With such high overall discrimination, just two or three erratic responses can drive the sensitive outfit statistics quite high. For example, the high outfit of *Few kinds of low-cost food for children* is driven by just three responses of households that denied this item while affirming 6 or 7 other items. Similarly, the high outfit of *Child did not eat whole day* is entirely due to two households that affirmed this item while denying all other child items. These few erratic responses out of a total of more than 3,000 in the estimation sample could indicate problematic items, but they may also just represent unusual incidents of inattention by a respondent or interviewer. In any case, their practical effects on prevalence estimates are very small.

Effects of Misfitting Items on Prevalence Rates

Returning now to the household scale (combined adult and child items), the practical effects on estimated prevalence rates of the two misfitting trichotomous items, *Adult cut size of meal or skipped meal* and *Adult did not eat for whole day,* are examined. The application of the food security measure that has the greatest public visibility and policy importance is the estimation of national and State prevalence rates of food insecurity and very low food security. The effects of item misfit on measured prevalence rates, then, provide important perspective on how seriously food security measurement is distorted by the misfit of these two items.

In spite of the questionable fit of these two items by conventional criteria for item-fit statistics, the effects of their misfit on prevalence rates is small. Table 2-9 compares observed and expected responses in level 2 (occurrence in more than 1 or 2 months) of the two items. These are evaluated at raw scores—omitting the upper level of the item under assessment—of 2 and 7

Table 2-9

Bias on measured prevalence rates due to misfit of *Adult cut size of meal or skipped meal* and *Adult did not eat for whole day* (average 2003-05)

	Food insecurity	Very low food security
Raw score, omitting the upper threshold of the variable in question[1]	2	7
Adult cut size of meal or skipped meal		
Percentage of households with child that have this raw score, omitting from the raw score calculation the upper level of Adult cut size of meal or skipped meal	4.485	1.376
Percentage of households in raw score observed in level 2	2.029	50.909
Percentage of households in raw score expected in level 2	.769	57.664
Bias on measured prevalence due to misfit[2]		
Percentage of households within the raw score	1.260	-6.755
Percentage of households with child[3]	.057	-.093
Percentage of all households[4]	.020	-.033
Adult did not eat for whole day		
Percentage of households with child that have this raw score, omitting from the raw score calculation the upper level of Adult cut size of meal or skipped meal	NA	1.346
Percentage of households in raw score observed in level 2	NA	2.558
Percentage of households in raw score expected in level 2	NA	1.724
Bias on measured prevalence due to misfit[2]		
Percentage of households within the raw score	NA	.834
Percentage of households with child[3]	NA	.011
Percentage of all households[4]	NA	.004

[1]The threshold for food insecurity is between raw scores 2 and 3. The threshold for very low food security is between 7 and 8.

[2]More responses than expected in level 2 biases the prevalence rate upward, since the raw score with the item in question omitted is just below the threshold for food insecurity or very low food security.

[3]Bias as a percentage of households with child is calculated as the product of the bias within the raw score group multiplied by the proportion of households with children that have that raw score.

[4]Bias as a percentage of all households is calculated as the product of the bias for households with children and the proportion of all households that have children (0.3521 in 2003-05).

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements.

(just below the thresholds for low and very low food security). It was noted that these two items would fit the Rasch model well if they were modeled as dichotomies, ignoring the frequency of occurrence. So the effect of misfit at the more severe level is the primary concern. Bias on the estimated prevalence of food insecurity (including low and very low food security) due to misfit of *Adult cut size of meal or skipped meal* amounted to +0.057 percent of households with children and +0.020 percent of all households. Since the estimated prevalence rate was 16.3 percent for households with children and 11.4 percent for all households, the bias due to misfit of this item was practically negligible. Bias on the estimated prevalence of very low food security was -0.093 percent of households with children and -0.033 percent of all households. Estimated prevalence rates were 3.9 percent for households with children and 3.7 percent for all households, so bias due to the misfit of the item was slight. The extent of bias due to misfit of *Adult did not eat for whole day* was effectively zero for the prevalence of food insecurity and negligible for the prevalence of very low food security.

These assessments do not reflect the full extent of the effects of misfit of these items on food security measurement using the continuous scale. Those effects are more substantial in the more severe range and might affect research applications of the measure. Nevertheless, the very small extent of the distortion in the primary monitoring applications suggests that changes in the way these items are incorporated in the current measurement methods need not be made precipitously.

Discussion

Household food security classification based on trichotomous modeling of the three dependent item sets is identical to that produced by the current standard methods, and continuous measures of household food insecurity differ only negligibly between the two methods. It is important, however, to model the items as trichotomies to obtain accurate assessments of how well they fit the Rasch model and whether, therefore, the use of raw score as an ordinal measure of food insecurity is justified.

The trichotomous items fit Rasch-model assumptions reasonably well in the adult scale and the child scale. The items also fit these models well if the followup items are omitted and the base items modeled as dichotomies. Decisions about whether to omit or include the frequency-of-occurrence followup items in these scales may, therefore, be made on the basis of either theoretical or practical considerations with little concern about the statistical integrity of the measure. The theoretical basis for keeping or dropping them depends in part on the methods considered most appropriate for incorporating frequency and duration of food insecurity into the measurement process. This topic was considered in a broader framework in Chapter 1.

In the household scale, which includes both adult and child items, the two adult-referenced trichotomous items *Adult cut size of meal or skipped meal* and *Adult did not eat for whole day* do not fit the model well, although the extent to which they distort measurement is slight. The household scale has other weaknesses—more serious than the somewhat problematic fit of these trichotomous items—that call into question its continued use as the primary food security measure for monitoring and research. Most notable are the nonequivalence of the thresholds for food insecurity in households with and without children and the bidimensionality between child-referenced and adult-referenced items, associated primarily with the age of the oldest child in the household. If the Federal food security measurement project decides to replace the main functions of the household scale with separate adult and child scales (and in some applications, a cross-tabulation of the two), then the problematic fit of the trichotomous adult items in the household scale will become moot.

However, if the household scale continues in use as the primary food security monitoring tool, then changes may be appropriate to avoid measurement distortions resulting from the misfit of the two trichotomous adult items. The effect of the misfit of the trichotomous items on estimated prevalence rates appears to be very small, so changes may be made with due deliberation, without concern that the current method is substantially distorting the main monitoring function of the measure. Or it may be deemed preferable to accept the relatively small distortions in order to preserve continuity of methodology and comparability of the statistical series of prevalence rates.

One way to deal with the misfitting item A*dult did not eat for whole day* would be to omit the frequency-of-occurrence followup from the scale. This would have minimal effects on measured prevalence rates of food insecurity and very low food security. These effects were calculated by observing the number of responses in level 2 of this trichotomy by households with raw scores in the first raw score above each threshold. During the period 2003-05, omitting the contribution of this item would have had no effect at all on the measured prevalence of food insecurity and would have reduced the estimated prevalence of very low food security by 0.033 percentage points (analysis not shown). Furthermore, detailed analysis (not shown) of observed-versus-expected responses to the other items in the scale indicate that most of those households should be classified as having low, rather than very low food security. The current method misclassifies them because of the weak fit of the trichotomous *Adult did not eat for whole day* item.

A practical consideration might suggest omitting this frequency-of-occurrence followup item since it contributes so minimally to measurement. Dropping the item would reduce respondent burden in most surveys. This would not affect respondent burden in the CPS-FSS because all of the questions about more severe food-insecure conditions include frequency-of-occurrence followup questions. These questions are used to characterize temporal patterns of food insecurity.

The followup to *Adult cut size of meal or skipped meal* plays a key role in the current food security measure since its severity is very near that of the threshold for very low food security. If that followup is to be omitted, it may be desirable to replace it with an item of similar severity in order to minimize disruption to the historical series. One possibility may be to split the base question into its two components, cutting the size of meals and skipping meals.

The trichotomous child item, *Child skipped meal*, fits the model reasonably well both in the household scale and the child scale. The practical consideration of reducing respondent burden might suggest omitting it. The effects of doing so would be negligible for the household scale and small, but not negligible for the child scale. On average over the period 2003-05, omitting this followup would have had no effect on the measured prevalence of food insecurity among children (raw score of 2 or greater on the child scale—a statistic not currently reported by USDA), but would have depressed the measured prevalence of very low food security among children (as measured by the child scale) by 0.063 percent of households with children, or 10.6 percent of households with very low food security among children.

A possible reason to retain the frequency-of-occurrence of *Adult did not eat for whole day and Child skipped meal* is that they contribute to identifying households with very severe levels of food insecurity. For some applications in very vulnerable populations, it may be important to differentiate these very severe levels. The items could be omitted from most surveys, and from the standard monitoring measure with minimal effect, and included as options in other surveys if additional reliability is needed in the severe range. If this approach is adopted, these items should be used only in the adult scale and the child scale. It is precisely in the more severe range of food insecurity where the misfit of *Adult did not eat for whole day* is likely to substantially distort measurement in the household scale.

Chapter 3

Measuring Food Security With a 2-Parameter Logistic Model

Abstract

Important characteristics of the U.S. Household Food Security Scale depend on the Rasch measurement model assumption that all items in the scale discriminate equally well. If this assumption is violated, then raw score may not be an ordinal measure of the latent trait, and measurement error could be reduced by using a more flexible 2-parameter measurement model that takes account of differences in discrimination across items. Nationally representative food security data are analyzed using both the single-parameter model and a 2-parameter model that relaxes the constraint of equal item discrimination. Comparisons of household measures based on the two scales indicate that use of a 2-parameter measure would result in little or no improvement in precision or reliability of food security measurement in the United States.

Background

The Committee on National Statistics (CNSTAT) panel that reviewed USDA's food security measurement methods recommended that USDA consider several "more flexible alternatives to the dichotomous Rasch model that underlies the current food insecurity classification scheme" (NRC, 2006). One of the alternatives identified by the panel, "Allowing the item discrimination parameters to differ from item to item when indicated by relevant data," is explored in this chapter.

The 1-parameter logistic (1PL or Rasch) model, which provides the statistical framework for the current measurement method, has several properties that make it particularly attractive for measuring food security:

1. Raw score is a consistent, ordinal measure of the underlying latent trait (food insecurity) under Rasch model assumptions. This makes it relatively easy to use the food security measure in applied settings. When item and measurement characteristics are known from national-level research, no specialized software or measurement statistics background are required to implement the measure in the field. Practitioners can simply add up the number of affirmative responses to determine each household's food security status.
2. The measure is relatively easy to explain to policy officials and the general public. Conditions in households at any specific measured level of severity can be explained in terms of the behaviors and experiences that are typically reported and denied by households at that level of food insecurity.
3. Provided that response data meet model assumptions, properties of the 1PL measure (i.e., the severity of each item and the respondent level on the latent trait corresponding to each raw score) can be estimated by methods that do not depend on modeling the distribution of the latent trait. Since a substantial proportion (usually a majority) of

respondents to food security surveys have raw score zero, it is often difficult to reliably characterize the distribution of the latent trait.

However, these desirable properties can only be assured if the response data are consistent with certain assumptions of the Rasch model, one of which is that the all items in the scale discriminate equally well. (This is conceptually equivalent to assuming that, in a linear factor analytic measurement model, all items have the same factor loading.) Items are assumed to differ only in severity—represented by a single parameter for each item. If response data do not meet the assumption of equal item discrimination, then raw score is not a sufficient statistic for food insecurity, and classification of household based on raw score could misrepresent the food security status of some households, In this case, items that discriminate more sharply (i.e., that are more strongly associated with the underlying latent trait) should be weighted more heavily when calculating the household's food insecurity. The correct weight for each item is its second parameter, and a 2-parameter scale is a weighted sum of affirmative responses to a set of items, with the weight of each item equal to its discrimination parameter.

The CNSTAT panel recommended that USDA explore whether, in fact, food security data fit the Rasch model assumption of equal item discrimination well enough to justify using a food security classification scheme based on the 1PL model. A 2-parameter logistic (2PL) model lacks the three desirable properties of the 1PL model listed above, so USDA prefers to use the 1PL model unless substantively important distortions in the measurement process are produced.

In this chapter, food security measures based on the 1PL model are compared with those based on a 2PL model. Using the nationally representative CPS-FSS data, correlations between household severity measures based on 1PL and 2PL models are examined, measurement errors between the two models are compared, and extent to which the two measures might differ in assigning food security status to households either discretely or probabilistically is examined.

These analyses answer both practical and theoretical questions about the measure. On the practical side, the two primary uses of the food security measure are:

- to categorize households by their food security status, for monitoring of prevalence rates of food insecurity nationally and for demographic and economic subpopulations.
- for research purposes, to examine the causes and consequences of food insecurity.

If the food security status of a substantial proportion of households is assigned differently by the two methods, then comparisons over time and across subpopulations could be affected. If such differences are rare, then the simpler 1PL model is preferred. If the measures are not highly correlated, then research results might differ (and would, presumably, be stronger with the more precise 2PL measure). However, if the measures are very highly correlated, then research results will not be degraded appreciably by use of the simpler model.

From a theoretical perspective, the issue of whether to use a simpler (1PL) or more complex (2PL) model depends on the extent of information lost by imposing the additional restriction of equal discrimination. A 2PL model is almost certain to fit the data better than a 1PL model, and in data sets as large as the CPS-FSS, the difference is almost certain to be statistically significant. However, even if a 2PL model does fit the data better, raw score provides a statistically consistent estimator for the severity of food insecurity provided the data meet other conditions on which the suitability of both methods depend.[1] The choice between the two models depends on how much information about the latent trait is lost by using the simpler model. The extent to which the correlation between the two measures falls short of unity and the extent to which measurement error of households differs between the two measures are indicators of the information in the 2PL model that is lost by using the simpler 1PL model.

[1]Raw score is a consistent estimator of the latent trait for data that fit any model that is: (1) monotone, (2) locally independent, and (3) unidimensional.

Data and Methods

Data

The main analyses used low-income subsamples of the nationally representative CPS-FSS data for 1995, 2002, 2003 and 2004. The 3 later years were analyzed jointly as representative of data collected using the current data collection protocol. In this protocol, some questions are not administered if responses to earlier questions all indicate high food security. The screens at data collection cause negligible distortions in food security prevalence estimates, but they distort the fit of IRT models somewhat by obviating a small proportion of highly improbable response patterns that might otherwise be observed. The 1995 data were collected under an earlier protocol in which all questions were administered to low-income households. Thus, they provide a check on the extent to which the 2002-04 findings might be affected by the screens.[2]

[2]It is possible, in principle, to take the screening into account in fitting IRT models. However, assessment of the 1995 data as collected and (in separate analysis) edited to appear as if they were collected under the later screening protocol indicated that the effects of the screening on estimated parameters and fit statistics were very small. Accordingly, I have applied both 2-parameter and 1-parameter estimation methods to the 2002-04 data with no adjustment for screening, treating the not-presented items as negative responses.

Data from the December 2010 CPS-FSS were used to assess the extent to which measurement error in prevalence rates would be reduced by using a 2PL measure rather than the 1PL measure now used by USDA. The 2010 data were used for this analysis because 2010 was the first year for which the Census Bureau calculated replicate weights for the Food Security Supplement to support calculation of sampling errors using balanced repeated replication (BRR) methods.

The samples were restricted to households with incomes less than 185 percent of the poverty line to avoid distortions that could arise from additional screening procedures that are applied to higher income households. Higher income households are screened out of the food security questions entirely unless they give an indication of some level of food-access difficulties on either of two preliminary screeners. Households with missing responses to one or more food security questions (a small proportion of households) were omitted. The resulting sample consisted of 14,387 households in 1995, 35,161 households in 2002-04, and 11,315 households in 2010.

Three scales were analyzed for each time period:

- Household scale of 15 items including 8 adult and household items and 7 child-specific items. This approximates the standard measure used by USDA in its annual food security reports. The full scale of 18 items includes how often did that happen followups to 3 of the 15 questions. The issue of whether to include these items in the scale and how to model them is the subject of Chapter 2 in this report. In order not to conflate the two issues, the followup questions were omitted from analyses in the present study. The samples for analyzing the household scale were limited to households with children present.
- Adult scale comprising the eight household and adult items. This approximates the 10-item adult scale that is used extensively for research purposes, although not in USDA's monitoring reports. The adult scale is preferred for many research applications because its measurement properties are thought to be unaffected or minimally affected by the presence or age of children. The samples for analyzing the adult scale included all households whether or not children were present.
- Child scale, comprising the seven child-specific items. This approximates the 8-item child scale used by USDA in its annual food security reports to further characterize households with children. The samples for analyzing the child scale were limited to households with children present.

Estimating 2-Parameter Models

Estimating 2PL item-response-theory (IRT) models is complicated by the fact that neither conditional maximum likelihood (CML) nor joint (unconditional) maximum likelihood (JML) solutions exist for typical response data (Fischer and Molenaar, 1995).[3] Marginal maximum likelihood (MML) estimation was, therefore, the preferred method and was implemented using Bilog 3 (Mislevy and Bock, 1990).

MML estimation is somewhat sensitive to the distribution of the latent trait in the estimation sample. This problem may be more salient in food security measurement applications than in other applications. Educational testing and similar applications and other applications typically involve a relatively large number of items that span the entire range of the latent trait among the respondents. Thus, almost all respondents are in the measurable range and the latent trait is measured with relatively high precision. As a result, the distribution of the latent trait in such data may be readily approximated. Food security survey data, on the other hand, typically include only a small number of items (rarely more than 15, often 10 or fewer), with the result that measurement of the latent trait is not very precise. Furthermore, the measurable range of food insecurity includes only a small proportion of the population; a large proportion of responses (typically up to 80 percent) have raw score zero. Approximating the distribution of the latent trait may, therefore, be difficult, and some distortion of the measure may result if the distribution is not modeled correctly.

The main 2PL (MML) analyses were based on a normal distribution prior for the latent trait. As a partial check on the robustness of the results to this specification, the analyses were repeated with an empirical prior specified for the

[3]Winsteps, which estimates a single-parameter model using JML methods, provides an item statistic labeled "discrimination" in its item statistics table. This is, however, only a first-order approximation based on item parameters that are estimated under single-parameter assumptions. This methodology was not considered to be an adequately rigorous basis for constructing a measure with the public and policy prominence of the Household Food Security Scale.

latent trait in the item estimation phase. This procedure allows the software to allocate the sample across a fixed number of quadrature points iteratively with estimation of item parameters so as to maximize model fit. Although item discrimination parameters were affected to some extent by allowing these more flexible distributions, the main results (ordering of household severity consistent with raw score and correlations of household measures with CML-based measures) were negligibly affected.

Household severity parameters and errors were the expected a posteriori (EAP) estimates based on a normal prior for the latent trait. Main analyses were replicated with EAP estimates based on an empirical prior and with maximum likelihood (ML) estimates; results did not differ substantially from the EAP-with-normal-prior results reported.

Estimating Single-Parameter Models

Item parameters and item-fit statistics based on the 1PL model were estimated using CML, following procedures described by Fischer and Molenaar (1995) and implemented in SAS data steps.[4] Household severities corresponding to each raw score were then estimated using maximum likelihood methods, also implemented in SAS data steps.

[4]The CML estimation programs were developed by the author and have been found to give identical results when tested against the commercially available Winmira software and against SAS Proc Logistic with the "Strata" option.

A 1PL model for the adult 2002-04 data was also estimated using JML implemented in Winsteps. These were included here for reference purposes. USDA does not currently use JML estimation for food security measurement, but initial scale development was based on JML analysis, and some food security researchers outside of USDA continue to use it. The main analyses in the current study compared 1PL models estimated in CML with 2PL models estimated in MML.

Household severity parameters were estimated using maximum likelihood, and errors were calculated as the inverse of the Fisher information function.

The CML and JML item and household parameter estimates were compared with corresponding MML statistics to examine the extent to which the three methods provide similar pictures of the structure of the measure.

Assessing Effects of Unequal Item Discrimination on Food Security Prevalence Estimates Based on Discrete Assignment of Food Security Status

The primary application of the food security scale in the realm of public statistics is to classify households with regard to their food security status. The food security status of the population and of selected subpopulations is then summarized by the prevalence of food insecurity at various levels of severity as represented by those classifications (see, for example, Coleman-Jensen et al., 2011). Two sets of analyses were conducted to assess the extent to which these classifications may be distorted by ignoring the unequal discrimination of items. The first set of analyses examined the differences in discrete assignment of food security status between the 1- and 2PL measures. The second set examined differences between the two measures in the probabilistic classification of food security status in which a probability of food insecurity is assigned to each household.[5]

[5]Alternatively, the probability of food insecurity for the interviewed household may be viewed as the estimated proportion of households in the population represented by that interviewed household that are food insecure. The CNSTAT panel that reviewed the food security measurement methods and recommended that USDA explore a 2-parameter measurement model also recommended exploring probabilistic assignment of food security status. Probabilistic assignment is assessed more fully in Chapters 5 and 6 but is examined in combination with the 2-parameter measure here because the 2-parameter measure might affect probabilistic assignment of food security status to a greater extent than discrete assignment.

To assess the extent to which discrete classification may be distorted by ignoring the unequal discrimination of items, the proportions of household pairs ordered differently by raw score and by the 2PL estimate of the latent trait were calculated in each sample. In these analyses, household pairs that differed in raw score were classified as either concordant (the 2PL measure of the latent trait was higher in the household with the higher raw score) or discordant (the 2PL measure of the latent trait was lower in the household with the higher raw score). The weighted sum of discordant pairs as a proportion of concordant plus discordant pairs was calculated as a measure of the extent of difference in ordering. This statistic is somewhat larger than the potential effect on food insecurity prevalence estimates resulting from use of the 1PL model because a substantial proportion of household pairs (typically 10 to 20 percent) have the same raw score, and some discordant pairs are at raw score breaks that do not correspond to a food security threshold (e.g., between raw scores 3 and 4).

Assessing the Effects of Unequal Item Discrimination on Interval-Level Measures of Food Insecurity Used in Research Applications

Some research applications use the household parameter—the value of the latent trait—in multivariate models. To assess the extent to which these analyses may be distorted by the 1PL measure, the Pearson correlations between the 2PL and 1PL household measures were calculated for each scale across the nonextreme cases in each sample.

Assessing the Effects of Unequal Item Discrimination on Food Security Prevalence Estimates Based on Probabilistic Assessment of Food Security Status

Mean errors for each measure were calculated across the nonextreme cases in each sample. In order to fairly compare errors across the two measures, two methods were used to equate the metrics of the 2PL and 1PL measures. The first method simply multiplied mean error by the average item slope, thus placing it on an approximately logistic metric. The second method multiplied mean error by the ratio of the standard deviation of item parameters in the 1PL model to the standard deviation of the item-severity parameters in the 2PL model. It is not clear which method provides the most defensible comparison, but in the event, differences between them were not great.

Comparison of measurement reliability (the proportion of total variance in the latent trait accounted for by the measure) overcomes the ambiguity in equating the 1PL and 2PL metrics, because reliability is invariant to linear transformations of the parameters. Reliability measures are calculated by decomposing the sum of squared differences of true household values on the latent trait from the grand mean. For the 1PL model, reliability (sometimes called Rasch reliability) is the weighted sum of squares between raw scores as a proportion of that sum plus the weighted sum of squared measurement errors. For the 2PL model, reliability is the weighted sum of squares between unique response patterns as a proportion of that sum plus the weighted sum of squared measurement errors,

The numerator in the 2PL calculation of reliability was also decomposed to the sum of squares between raw scores and the sum of squares between unique response patterns within raw scores. The relative size of these sums of squares provides insight into the extent to which taking account of the difference in item discrimination reduces measurement error.

The correlations between 1PL and 2PL measures of the probability of food insecurity (and, separately, of very low food security) were calculated to assess the extent to which probabilistic assignment of food security status may differ between the two measures. For these analyses, the probabilities of food insecurity on each measure were calculated as follows:

- A threshold value on the latent trait was selected so as to equate the mean probability of food insecurity across all households to the proportion classified as food insecure, based on discrete assignment by raw score.
- The probability of food insecurity for each household for each measurement method was calculated as the proportion of a normal distribution beyond the threshold, given a mean of the distribution equal to the household's severity-parameter estimate and standard deviation equal to the measurement error.
- The population estimate was calculated as the weighted mean of probabilities across households.[6]

This calculation was repeated for food insecurity and very low food security for each measure.

In the calculations of equivalent thresholds, households with zero raw score were included and considered to be fully food secure with no measurement error, and households with maximum raw score were included and considered to have very low food security with no measurement error. However, households with these extreme scores were omitted from the correlation calculations. Measurement error cannot be calculated for extreme scores without making assumptions about the distribution of the true latent trait. This is especially true for households with raw score zero, which comprise 40 to 60 percent of the households in these samples.

Assessing the Effects of Unequal Item Discrimination on the Precision of Food Security Prevalence Estimates

The bottom-line assessment of the 2PL model is the extent to which it would improve the precision of USDA's annual prevalence estimates of food insecurity and very low food security in U.S. households. Replication-based estimates of errors in prevalence statistics directly assess the most important practical effects of differences in measurement error between the 2PL and 1PL measures.

Errors of estimated prevalence rates of food insecurity and very low food security based on discrete assignment of food security status and on probabilistic assignment by 2PL and 1PL measures were calculated using two replication methods. Jackknife replication methods were used for the 2002-04 data and balanced repeated replication (BRR) methods for the 2010 data. Both of these methods are designed for the primary purpose of estimating sampling error, but in fact, both methods estimate the combined sampling error and measurement error. Indeed, the two sources of error cannot be

[6]In practice, an approximate threshold was estimated initially and then adjusted iteratively until the weighted probability of food insecurity in the sample matched the target proportion—that based on discrete assignment of food security status.

differentiated by these methods. In each analysis in this study, sampling error was the same for the two measurement methods, since they were based on the same data, so differences in Jackknife- or BRR-estimated errors can be ascribed to differences in measurement error.

The jackknife estimates used the eight month-in-sample "rotation" groups in each year, since these are, effectively, independent samples of U.S. households. Standard jackknife replication methods were applied to these subsamples to calculate error estimates for prevalence rates of food insecurity and very low food security for each of the three scales for the 3-year period 2002-04.

The BRR methods using the December 2010 CPS-FSS data were implemented in the SAS SurveyMeans procedure using replicate weights provided by the Census Bureau.[7]

Findings

The three estimation methods provide consistent pictures of the structure of the food security measure. In the 2002-04 data, severity parameter estimates for the adult scale based on an assumption of equal item discrimination, whether estimated using JML or CML methods, were distorted only slightly by the differences in discrimination across items (table 3-1 and fig. 3-1). In the 2PL model, none of item discriminations diverged greatly from unity.

[7]Replicate weights were first provided for the CPS-FSS in 2010. The weights are based on the Fay method and, following specifications provided by the Bureau, the option "Fay=.5" was specified in the SAS SurveyMeans procedure.

Table 3-1

Item severity and discrimination parameters estimated from nationally representative CPS-FSS data using MML methods, and comparison to item severity parameters estimated using single-parameter JML and CML methods

	Item severity					
Item	MML 2-param.[1]	JML 1-param.[2]	Diff. JML-MML	CML 1-param.[2]	Diff. -CML MML	Item discrimination[1] (MML)
Worried food would run out	-3.65	-3.77	-0.12	-3.68	-0.03	1.00
Food bought did not last	-2.62	-2.52	.10	-2.55	.06	.97
Could not afford balanced meals	-1.84	-1.94	-.10	-1.99	-.15	.64
Adult cut size of meals or skipped meals	-.11	-.03	.08	-.07	.04	1.15
Ate less than felt should	-.13	-.01	.12	-.04	.08	1.25
Hungry but did not eat	1.83	1.89	.06	1.93	.10	1.08
Lost weight	2.80	2.78	-.02	2.83	.02	1.00
Adult did not eat for whole day	3.71	3.59	-.12	3.58	-.13	.92
Mean	0.00	0.00		0.00		1.00
Standard deviation	2.63	2.63		2.63		
Discrimination relative to that of the MML 2-parameter model		1.17		0.93		

JML = Joint (or unconditional) maximum likelihood.
CML = Conditional maximum likelihood.
MML = Marginal maximum likelihood.

[1]Estimation of the MML 2-parameter model was constrained to have mean item slope of 1.0 and mean item severity parameter of 0.

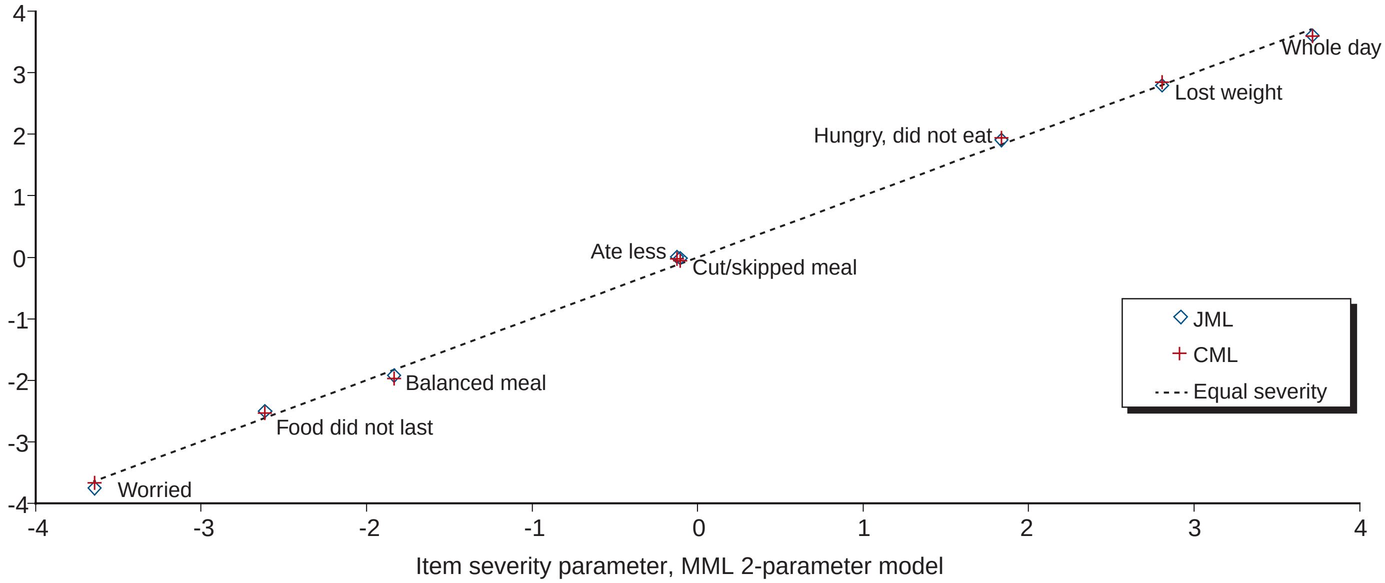
[2]JML and CML item-severity parameter estimates were adjusted by a linear transformation so that the mean and standard deviation of the estimated severity parameters matched the mean and standard deviation of the MML estimates. The ratio of standard deviations of item scores prior to the transformation is the item discrimination compared with that of the MML model as entered in the bottom row of the table.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements (CPS-FSS) of 2002, 2003, 2004 for households with incomes less than 185 percent of the Federal poverty line.

Figure 3-1

Single-parameter JML and CML item severity parameter estimates compared with MML 2-parameter estimates

Item severity parameter, JML and CML 1-parameter models



Notes:
CML = Conditional maximum likelihood.
JML = Joint (unconditional) maximum likelihood.
MML = Marginal maximum likelihood.

Estimation of the MML 2-parameter model was constrained to have mean item slope of 1.0 and mean item- severity parameter of zero. JML and CML item-severity parameter estimates were adjusted by a linear transformation so that the mean and standard deviation of the estimated severity parameters matched the mean and standard deviation of the MML estimates.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2002, 2003, 2004 for households with incomes less than 185 percent of the Federal poverty line.

The discrimination parameter of *Could not afford balanced meals* was somewhat low (0.64) and that of *Ate less than felt should* somewhat high (1.25). Discrimination parameters for the remaining items ranged only from 0.92 to 1.15.

The divergences from the Rasch-model assumption of equal discrimination, as estimated using the 2PL MML method, are partially reflected by the item-infit statistics in the 1PL models (table 3-2). Infit statistics for *Could not afford balanced meals* were 1.22 and 1.27, respectively, in the JML and CML models, while those for *Ate less than felt should* were 0.77 and 0.79, respectively.

Differences in discrimination of this magnitude are not likely to distort substantially measurement based on Rasch-model assumptions. The distribution of 2PL household severity estimates is likely to be heavily concentrated around the 1PL estimate within each raw-score group. This was confirmed by the correlation between household measures based on the three models. For households with nonextreme responses, the Pearson correlations between household measures based on the 1PL model (whether estimated by JML or CML) and those based on the 2PL model was .990. Very little information provided by the 2PL model is lost by imposing the equal-item-discrimination constraint in the 1PL model.

The ordering of households by raw score on the adult scale was also perfectly consistent with the order based on the 2PL model. The maximum value of the household measure in each raw score group was smaller than the minimum value in the next-higher raw score group (table 3-3). Thus, if classification of households by food security status is discrete based on their estimated severity on the latent trait, and if national prevalence rates are constrained to be equal on the two methods, then classification of every household by the two measures is identical.[8] Not only would prevalence rates of food insecurity and very low food security in all subpopulations be the same on both measures, but exactly the same households would be in each category on both measures. Similarly, prevalence rates across years would be the same based on the two measures, provided the combined multiyear prevalence rates were constrained to be equal for the two measures.

[8]This constraint provides a reasonable basis of comparison given the character of the food security status concepts. The selection of thresholds for defining categories of food insecurity is arbitrary over some range approximating the defining characteristics. To compare prevalence rates, then, we would select a threshold on the 2-parameter measure that gave the same prevalence as the raw-score-based threshold at the national level.

Table 3-2

Item-infit statistics calculated from nationally representative CPS-FSS data based on the 1-parameter (Rasch) model using JML and CML methods

Item	JML	CML
Worried food would run out	1.06	1.01
Food bought did not last	.88	.90
Could not afford balanced meals	1.22	1.27
Adult cut size of meals or skipped meals	.81	.83
Ate less than felt should	.77	.79
Hungry but did not eat	.88	.89
Lost weight	.95	.96
Adult did not eat for whole day	1.05	1.00

JML = Joint (or unconditional) maximum likelihood.
CML = Conditional maximum likelihood.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements (CPS-FSS) of 2002, 2003, 2004 for households with incomes less than 185 percent of the Federal poverty line.

Table 3-3

Household estimated measures of severity of food insecurity based on MML methods, by raw score on 8-item adult food security scale

	MML household measure		
Raw score	Minimum	Mean	Maximum
0	-8.33	-8.33	-8.33
1	-4.90	-4.22	-3.49
2	-2.90	-2.51	-1.82
3	-1.64	-1.44	-.62
4	-.59	-.18	.35
5	.38	.88	1.34
6	1.46	1.96	2.42
7	2.81	3.25	3.77
8	5.37	5.37	5.37

MML = Marginal maximum likelihood.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2002, 2003, 2004 for households with incomes less than 185 percent of the Federal poverty line.

Table 3-4

Extent of difference in household classification by raw score between 1-parameter (1PL) and 2-parameter (2PL) models, and correlation between household severity measures based on the two models

Scale	Households out of order on raw score[1]	Correlation between 2PL and 1PL severity measures[2]
	Percent	*Pearson correlation (N)*
Adult 8-item scale		
2002-04	0.00	.990 (14,678)
1995 (all items presented)	.00	.995 (5,644)
Household 15-item scale		
2002-04	.00	.999 (8.278)
1995 (all items presented)	<.001	.998 (3,408)
Child 7-item scale		
2002-04	.12	.991 (5,495)
1995 (all items presented)	.03	.995 (2,063)

Analysis samples were restricted to households with incomes less than 185 percent of the poverty line and with no missing responses to scale questions presented. Household- and child-scale samples were further restricted to households with children. All statistics were calculated using sampling weights; however, statistics based on unweighted cases differ negligibly if at all.

[1]Household pairs for which the 2PL measure of severity was inconsistent with severity order based on raw score, expressed as a percentage of household pairs with different raw scores. Households with extreme raw scores (zero and maximum) were omitted from the analysis.

[2]Samples for correlation analyses were restricted to nonextreme responses.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2002, 2003, 2004 for households with incomes less than 185 percent of the Federal poverty line.

Results for the adult scale were similar when based on the 1995 data. Again, no households were out of order by raw score alone, and the inter-measure correlation across households was .995 (table 3-4).

Results for the 15-item household scale and 7-item child scale were only slightly less consistent. In the 2002-04 data, the 1PL scale was perfectly ordered relative to the 2PL, and the inter-measure correlation was .999. A very small proportion of household pairs (less than 0.001 percent) were discordant in the 1995 data, and the inter-measure correlation across household was .998. Discordant household measures were somewhat more common in the child scale, although still only about one-tenth of 1 percent. Correlations between the two measures across households were .991 in the 2002-04 data and .995 in the 1995 data.

Analysis of measurement error indicates that the 2PL model also provides little or no improvement in probabilistic assignment of food security status. Mean measurement error on the adult and household scales was about the same for the two methods, or perhaps slightly smaller for the 1PL measure

Table 3-5

Characteristics of 1-parameter (1PL) and 2-parameter (2PL) measures of food insecurity, taking into account measurement error

Statistic	Adult 8-item scale	Household 15-item scale	Child 7-item scale
Ratio of mean measurement error, 2PL/1PL, with 2PL errors adjusted to logistic metric[1]	1.085	1.020	0.899
Mean measurement error, 2PL/1PL, with 2PL errors adjusted to equate standard deviation of item severity parameters to those of the 1PL model[2]	1.012	0.986	0.786
Rasch reliability, 2PL model	.756	.828	.701
Rasch reliability, 1PL model	.763	.829	.671
Decomposition of squared errors, 2PL model (proportion of total squared error)			
Between raw scores	.7431	.8265	.6909
Between unique response patterns within raw scores	.0127	.0015	.0105
Within unique response patterns	.2442	.1720	.2986
Correlation between 2PL and 1PL estimates of the probability that the household is food insecure[3]	.989	.999	.987
Correlation between 2PL and 1PL estimates of the probability that the household has very low food security (very low food security among children for the Child Scale)[3]	.996	.999	.996
Number of cases (unweighted)	14,678	8,278	5,495

Analysis samples were restricted to households with incomes less than 185 percent of the poverty line, with no missing responses to scale questions presented. Households with extreme responses (raw score zero or maximum) were omitted. Household- and child-scale samples were further restricted to households with children. All statistics were calculated using sampling weights. Household severity parameters for 1PL logistic models were maximum likelihood estimates based on item parameters estimated using conditional maximum likelihood methods. Errors for 1PL models were calculated as the square root of the inverse of the Fisher information function. Severity parameters and errors for the 2PL logistic models were the expected a posteriori (EAP) estimates.

[1]The 2PL error estimates were multiplied by the mean slope of the items to place them on a logistic metric.

[2]The 2PL error estimates were multiplied by the ratio of the standard deviation of item parameters, 1PL/2PL, to place the two scales on the same metric.

[3] Thresholds were selected for each measure so as to produce the same population-level prevalence rates as those given by discrete assignment (including extreme raw scores). Probability of food insecurity (or very low food security) for each household was then calculated as the proportion of a normal distribution that would fall beyond the threshold, given a mean at the severity parameter estimate for the household and a standard deviation equal to the estimated measurement error for the household. Households with extreme scores were omitted from the correlation calculations.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2002, 2003, 2004 for households with incomes less than 185 percent of the Federal poverty line.

(table 3-5). Rasch reliability, which avoids uncertainties of equating slopes on the two measures, was also essentially the same for the two measures, or slightly higher for the 1PL measure. For the child scale, however, the 2PL measure did reduce measurement error by 10 to 20 percent and increase Rasch reliability by about 4 percent.

The reason for the lack of substantial improvement due to the 2PL model is seen in the decomposition of the sum of squared differences from the grand mean. The total variance in the latent trait accounted for by the 2PL model is the sum of squares between raw scores and the sum of squares between unique response patterns within raw scores. The latter is a very small proportion of that total for all three scales.

The result of the similarities in error structure between the two measures is that the probabilities of food insecurity and very low food security are highly correlated between the two measures for all three scales. Pearson correlations between probabilities of food insecurity ranged from .987 to .999.

By far, the applications of the food security measures with the highest public visibility and greatest policy relevance are USDA's annual estimates of the prevalence rates of food insecurity and very low food security in U.S. households. Thus, the most relevant assessment of the 2PL measure, for practical purposes, is the extent to which it reduces measurement error on those key statistics. The findings in this regard provide no evidence that the 2PL measure would improve measurement. For all three scales in both time periods and error estimation methods, combined sampling and measurement error were essentially the same for probabilistic assignment of food security status by the two

Table 3-6

Combined measurement and sampling error of prevalence rates of food insecurity and very low food security calculated from discrete assignment of food security status based on raw score and from probabilistic assignment of food security status based on 1-parameter (IPL) and 2-parameter (2PL) measures of food insecurity, taking into account measurement error

Sample, severity level, and method	Combined measurement and sampling error		
	Adult 8-item scale	Household 15-item scale	Child 7-item scale
	Percentage points		
2002-04 food insecurity			
Discrete assignment based on raw score	0.389	0.682	0.539
Probabilistic based on 1PL measure	.342	.652	.510
Probabilistic based on 2PL measure	.347	.652	.511
2002-04 very low food security			
Discrete assignment based on raw score	.196	.279	.106
Probabilistic based on 1PL measure	.182	.259	.073
Probabilistic based on 2PL measure	.185	.260	.077
2010 food insecurity			
Discrete assignment based on raw score	.493	.817	.660
Probabilistic based on 1PL measure	.440	.722	.530
Probabilistic based on 2PL measure	.446	.724	.530
2010 very low food security			
Discrete assignment based on raw score	.322	.448	.211
Probabilistic based on 1PL measure	.265	.382	.153
Probabilistic based on 2PL measure	.266	.382	.154

Analysis samples comprised households with incomes less than 185 percent of the poverty line and with no missing responses to scale questions presented. Household- and child-scale samples were further restricted to households with children. Households with raw score zero were assumed to be fully food secure with no measurement error. Households with maximum raw score were assumed to have very low food security with no measurement error. All statistics were calculated using sampling weights.

Household severity parameters for 1PL logistic models were maximum likelihood estimates based on item parameters estimated using conditional maximum likelihood methods. Errors for 1PL models were calculated as the square root of the inverse of the Fisher information function. Severity parameters and errors for the 2PL logistic models were the expected a posteriori (EAP) estimates.

Thresholds were selected for each measure so as to produce the same population-level prevalence rate as that given by discrete assignment. Probability of food insecurity (or very low food security) for each household was then calculated as the proportion of a normal distribution that would fall beyond the threshold, given a mean at the severity parameter estimate for the household and a standard deviation equal to the estimated measurement error for the household.

Combined measurement and sampling errors for the 2002-04 sample were estimated using Jackknife replication methods based on the eight month-in-sample groups, which are independent samples within each year. Errors for the 2010 sample were estimated using balanced repeated replication methods based on replicate weights provided by the Census Bureau for the 2010 data.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2002, 2003, 2004, 2010 for households with incomes less than 185 percent of the Federal poverty line.

measures (table 3-6). Where differences did exist, they were very small and in favor of the 1PL model. It is worthwhile noting that probabilistic assignment by either measure reduced measurement error compared with discrete assignment based on raw score (the current method) by 4 to 18 percent for the adult and household scales and by up to 31 percent for the child scale.

Conclusions

There is no evidence that food security measurement at the national level would be improved by basing it on a 2PL model rather than the 1PL (Rasch) model that is the current standard. Differences in household assessment and measurement error between the two measures are negligible. Errors on estimates of national prevalence rates are essentially the same for both methods. To a great extent, the lack of difference is a result of careful work done in the original development and selection of the items that make up the scale. Items that did not fit the 1PL model reasonably well were omitted from the scale (Hamilton et al., 1997b).

The 1PL Rasch model is, therefore, preferred because of its simplicity of implementation in the field; its transparency, which makes it relatively easy to explain to policy officials and the public; and its greater statistical simplicity. Two-parameter analysis may contribute to item and scale assessment if the performance of an item appears to be poor in a specific subpopulation, or to deteriorate in the future.

Chapter 4

Assessing Differential Item Function in the U.S. Household Food Security Scale

Abstract

This chapter assesses the recommendation by the Committee on National Statistics (CNSTAT) that USDA "… [fit] models that allow for different item parameters for households with and without children for the questions that are appropriate for all households in order to study the possibility and effects of differential item functioning." The recommended assessment was conducted along with assessment of differential item function for several other subpopulations of interest, using data from 11 administrations of the Current Population Survey Food Security Supplement.

The key measurement characteristics of items in the adult food security scale were found to be generally similar across subpopulations that differ by age, sex, race, household composition, Hispanic ethnicity, and region. Differential item function (DIF) was found between households with and without children and was found to be somewhat greater for Hispanic households with children. DIF was substantial for only one item, We couldn't afford to eat balanced meals. However, the practical implications of the DIF are modest, and correcting for them would not change the assessment of intergroup differences in food security in any important ways.

Background

The Committee on National Statistics (CNSTAT) panel that reviewed USDA's food security measurement methods recommended that USDA consider several "more flexible alternatives to the dichotomous Rasch model that underlies the current food insecurity classification scheme" (NRC, 2006). One such alternative, "… [fitting] models that allow for different item parameters for households with and without children for the questions that are appropriate for all households in order to study the possibility and effects of differential item functioning," is explored in this chapter.

The multiple-item Rasch model, on which the U.S. Household Food Security Scale is based, assumes that the severity and discrimination of items are invariant across subpopulations. Violation of this assumption between two subpopulations is described as differential item function. The underlying cause of DIF could be either that respondents in two subpopulations understand the question to refer to different objective conditions, or that the way the two subpopulations experience or manage food insecurity differs so that the relative severity of items differs between the subpopulations. The practical effect of DIF is that severity of food insecurity at a given raw score may not be the same in the two subpopulations, and prevalence statistics based on classification by raw score may give a biased comparison between the two subpopulations.

The specific DIF raised as a concern by the CNSTAT panel was DIF between households with and without children. This may be of particular concern because equivalence between the 18-item scale applied to households with children and the 10-item scale applied to households without children is established based on the 10 items that are common to the two scales (Bickel et al., 2000). If 1 or more of those 10 items function substantially differently in households with and without children, then the standard classification methods may not provide unbiased comparisons of food security in the two subpopulations.

Although the CNSTAT panel specifically mentioned only DIF between households with and without children, I considered it worthwhile to test for DIF between several other population groups as well. Over the years, in various policy discussions, questions have been raised about comparability of food security measurement between Hispanics and non-Hispanics, rural and urban households, and male and female respondents. Examining item function between these and other groups provides insight into whether measured prevalence rates based on current methods give unbiased comparisons of food insecurity among these various groups.

Practical Considerations Relative to Pending Decisions About Other Measurement Issues

As discussed elsewhere in this report, nonequivalence of the threshold for food insecurity between the 18-item scale applied to households with children and the 10-item scale applied to households without children biases prevalence estimates of households with children upward substantially relative to those of households without children.[1] Preliminary analysis suggests that any bias due to DIF is small compared with this nonequivalent-threshold bias. The relevance and importance of any DIF that may exist depends on how USDA decides to solve the more fundamental classification issue. The three most likely scenarios have different implications for this question:

- If the current classification methods are continued, then any effect of DIF on measured food insecurity is likely to be small relative to the bias that exists due to the 18-item-versus-10-item-scale bias. In that case, DIF effects on scale equivalence would be of practical relevance mainly for the very low food security classification.
- If the 10-item adult scale (or of cross-classification by separate adult and child scales) replaces the 18-item scale as the main monitoring measure, then the effects of DIF between households with and without children would be of interest at both the food-insecure and very-low-food-security thresholds.
- Depending on the extent of bias due to DIF and to communication and credibility considerations, USDA may decide either to adjust prevalence rates of some subpopulations for estimated DIF-based bias or to apply a uniform measurement and classification system for all households but make note of any known substantial inter-group biases to facilitate appropriate interpretation.

Previous Research

The question of scale invariance between households with and without children was examined by the Abt Inc. team in its original work to develop the

[1]Households are classified as food insecure if they affirm three or more items, but this threshold represents a lower level of severity on the 18-item scale applied to households with children than on the 10-item scale applied to households without children. At the national level in recent years, the resulting bias has amounted to about 3 percentage points on a measured prevalence of 20 percent in households with children (Coleman-Jensen et al., 2011, p. 10).

food security scale. Hamilton et al. (1997b, p. 18) state (after describing development of the scale for households with children), "Invariance tests were also performed for households without children, subdividing them into households with any elderly members (age 60 or over) and household with no elderly members. …. In this procedure, we separately fit the model to each subpopulation, such as households with children, households with no children but with elderly members, and households with neither children nor elderly. Each of the separate models was then used to compute scale values [i.e., respondent measures] for all households in the full sample. The values computed with the different models were then compared through plotting and correlation analysis." Hamilton et al. (1997b, p. 18) described the results of these analyses as follows, "The model replications provided clear support for the invariance of the primary measurement model across subsamples, as well as across different types of households. In each replication, the item calibrations gave identical or near-identical rankings of item severity and consistent clustering of closely ranked items. Applying models fit on separate subsamples yielded household values that correlated at the .99 level."

Ohls et al. (2001, appendix E) revisited the issue of intergroup invariance of the food security measure in their analysis of the 1995-1997 Current Population Survey Food Security Supplement (CPS-FSS) data.

> "Here, we report findings from estimating models for subsets of households classified by 1) race/ethnicity; 2) household composition, including the presence of children; 3) metropolitan versus nonmetropolitan locations; and 4) region…. Our overall conclusions are that the results obtained from the Rasch model are reasonably robust when examined by subgroup."

The methods used by Ohls et al. did not include formal assessment of DIF. Nevertheless, the comparison of item severities for households with children versus all households (the majority of which do not have children) gives no indication of substantial DIF (Ohls et al., 2001, table E.4 and figure E.2).

Wilde (2004) compared response patterns to the household- and adult-referenced food security items by households with and without children. His reported finding of substantial differences in these response patterns may have motivated the CNSTAT recommendation to investigate this issue. However, the Wilde (2004) assessment was flawed methodologically. Models were estimated separately for households with and without children, but were constrained to have the same severity parameter for the least severe item (worried that food would run out). All other item parameters were then found to differ between households with and without children, and the differences were statistically significant.

However, my secondary analysis of the Wilde results revealed that if the two scales were adjusted to a common metric using the more conventional method of equating the mean of item parameters, almost all of the child/no-child differences became nonsignificant. It appears that the main difference was in the severity of the least severe item. The Wilde article did not estimate the practical implications of the differences in response patterns either for comparability of prevalence estimates based on the 10-item scale or for equating the 10-item and 18-item scale.[2]

[2]Unfortunately, the discussion of implications in Wilde (2004) did not separately assess the effect of DIF and the effect of the noncomparable raw-score thresholds between the 18-item and 10-item scales. As a result, it is not clear whether the extent of DIF reported was of substantive importance.

Opsomer et al. (2003) used generalized linear mixed models to simultaneously estimate levels of food insecurity and measurement model parameters using the 1995, 1996, and 1997 CPS-FSS. By interacting household characteristics with item responses, these models also provided comparisons of item characteristics across subgroups, including households with and without children. Although some differences were found (primarily among racial-ethnic minorities and between metropolitan and nonmetropolitan respondents—not between households with and without children), the practical implications of these differences were found to be small.

The present study adds to the earlier research in several ways (details are provided in the Data section):

- More extensive data are used than in the earlier studies, including CPS-FSS data aggregated from 11 annual surveys—more than 46,000 households with complete and nonextreme responses. This supports very precise parameter estimates and avoids spurious findings due to peculiarities of a single survey.
- The extent of bias on prevalence estimates between groups where DIF appears to be substantial enough to be of possible concern is assessed. This assesses the practical importance of any DIF that exists. In such large samples, there will be many statistically significant differences in item function between subpopulations. The practical importance of the differences depends not on their statistical significance, but on the extent to which they affect comparability of prevalence statistics across subpopulations.
- The two "how often did this happen?" followup questions and their base questions are modeled as trichotomies, rather than as if they were independent dichotomous items. Although these dependencies have only small effects on item-severity parameter estimates, taking account of the dependencies is technically correct and removes any distortions that may have resulted from ignoring their dependence in earlier analyses.
- Screening within the module is taken into account, and the data are re-edited to correct for different screening in households with and without children. These differences may have distorted comparisons in the Wilde (2004) analysis.

Data

Analyses were based on food security data collected in 11 administrations of the CPS-FSS from 1998 to 2007. (Two surveys were administered in 2001, one in April and one in December.) The analysis sample for assessing DIF was restricted to households with incomes less than 185 percent of the poverty line. Most households with incomes higher than 185 percent of the poverty line were screened out of the food security questions. Data for households at all income levels were used to calculate a reference distribution of households across raw score groups. This distribution was used to calculate the bias in prevalence estimates due to DIF, following procedures described below in the "Methods" section.

Rasch-model-based assessments were based on the 10-item Adult Food Security Scale, comprising the adult-referenced and general household items in the CPS-FSS. Households with missing responses to any of the food security questions (about 2 percent of the sample) were omitted from all analyses. Households with extreme raw scores (0 and 10) on the Adult Food Security Scale (59 percent of the sample) were omitted from the model-fitting analyses but were included in the reference distribution of households across raw score groups used in the calculation of bias in prevalence estimates due to DIF.

Rotation groups that were administered nonstandard food security scale questions were omitted from all analyses. In some years, split ballot tests were conducted for question-development purposes. Responses to these test questions would not be consistent with responses to the standard questions. Omitted groups were as follows:

- HRMIS=8 in 1998 and 1999. (Certain questions were referenced to the respondent, rather than to all adults in the household, or to an individual child, rather than to all children in the household.)
- HRMIS=8 in 2000. (Alternative forms of the three general household questions were tested for Health Canada.)
- HRMIS=3 or 8 in 2007. (Alternative form for balanced meals question was tested.)

Two adjustments were made to ensure that differential screening of households with and without children did not bias the comparisons of item parameters. At two points in the administration of the series of food security questions, responses are evaluated to determine if the remaining food security questions, will be administered or skipped (Bickel et al., 2000). Prior to 2007, child-referenced questions as well as adult-referenced and general household questions, were taken into consideration in assessing these screens. As a result, some households with children were administered adult-referenced questions that would not have been administered to households without children that gave the same responses to the questions in the adult scale. Since these differences could potentially distort comparisons of item-severity parameter estimates, responses of households with children were edited to reflect screening as it would have been implemented if there were no children in the household. Specifically, cases were omitted from all analyses if they responded "never" to the initial three food security questions (0.4 percent of in the complete, nonextreme analysis sample). Such households, if they had no children, would have been screened out of the remaining food security questions. They would, therefore, have had raw score zero and been omitted from the Rasch model estimation.[3]

The second adjustment for screening was to edit responses to the final two questions—about not eating for a whole day—to "no" for households with children that would have been screened out of those questions except for their response to the child item in the second block of questions. This adjustment affected only 18 cases (0.07 percent of the complete, nonextreme analysis sample of households with children.)

The final sample sizes for the Rasch-model analyses (those with complete and nonextreme responses) were 24,418 households with children and 22,546 households without children.

[3]Households either with or without children were also administered the second block of food security questions if their response to HH1 (a question not included in the scale) was "sometimes not enough to eat" or "often not enough to eat," even if they responded "never" to the first three questions in the scale. However, such households (0.4 percent of households with complete, nonextreme responses) were omitted from the analysis sample in this study because the estimation software, which takes account of screening within the module, can only handle data that are consistent with screens based on items in the scale.

Methods

Assessing Whether Any Items Exhibit Differential Function

Item-severity parameters were estimated separately for the child and no-child subsamples using SAS programs developed by ERS to implement conditional maximum likelihood (CML) estimation methods.[4] The programs treat the "how often did that happen?" followup questions along with their base questions as trichotomies and take account of the internal screens in the food security module.

As a first step in examining the items for possible DIF, item-severity parameters for households with no child were compared with those for households with children. The standard deviation of item parameters for households with no child was 9 percent smaller than that for households with children, indicating somewhat lower average item discrimination.[5] To place the scales for the two types of households on a common metric, therefore, parameters for households with no child were multiplied by a constant calculated to equate the standard deviation of item parameters to that of households with children. A constant was then added to equate the mean of the parameters for households with no child with that of households with children. The adjusted severity parameters for households with no child were then plotted against those for households with children. A signed area test (Raju, 1988) was also conducted in which the difference in probability of an affirmative response by households without and with children was weighted by the distribution of households with children across raw score groups.

Initial examination of the plot and the signed area test statistics suggested that one item, *Could not afford balanced meals*, had substantial DIF. Accordingly, a second adjustment (adjustment #2) was made to the metric of the scale for households with no child, omitting the item, *Could not afford balanced meals*, from the set of items for which mean and standard deviation were equated. The adjusted item parameters were again plotted against the item parameters for households with children, and a second signed area test was conducted based on the adjusted parameters.

DIF appeared to be relatively small for items other than *Could not afford balanced meals*. However, in the primary method used subsequently to estimate the extent of bias due to DIF, the remaining items were not constrained to be equal in the two samples. Only the mean and standard deviation of those items were constrained to be equal. Thus, the analysis allowed for DIF of the other items, but with the constraint that, on average, the items represented the same level of severity in the two subsamples.[6] In a confirmatory analysis, the average item discrimination and the severity parameters of all items other than *Could not afford balanced meals* were constrained to be equal in the two subsamples.

Exploring the Reason for DIF Between Households With and Without Children

It was not clear initially whether the DIF for the item *Could not afford balanced meals* between households with and without children resulted from the presence of children in the household per se or from other characteristics

[4]The SAS programs use Newton-Raphson iterative approaches implemented in a data step to estimate item parameters following methods described by Fisher and Molenaar (1995). The ERS programs have been tested against SAS Proc Logistic with the new "Strata" command in multiple data sets with dichotomous items and found to give identical results. They were also tested on simulated data with trichotomous items that were generated to be perfectly consistent with polytomous Rasch model assumptions and found to recover the generating parameters exactly and to give perfect item-fit statistics.

[5]For the two trichotomous items, all adjustments, comparisons, and presentations of item parameters used the "50 percent probability" point (sometimes called the "Rasch-Thurstone threshold") for the 0/1 and 1/2 thresholds. The 0/1 threshold is the severity level at which the probability of a response in the "0" category is .5. The 1/2 threshold is the severity level at which the probability of a response in the "2" category is .5. The Rasch-Thurstone thresholds are more intuitively meaningful than the pure Rasch thresholds, which are inverted for these items (i.e., the Rasch severity parameter for the 1/2 threshold is lower than that for the 0/1 threshold).

[6]This is characteristic of the problems of assessing DIF in scales based on a small number of items in cases where it cannot be assumed that any specific item or set of items functions identically in the two subpopulations. Equivalent function can only be inferred based on observed similarity of relative severity of items.

that differ systematically between households with and without children, such as sex, age, race, ethnicity, and other characteristics of the respondent. Respondents in households with children are more likely to be female, younger, and Black or Hispanic than respondents in no-child households. Education and economic circumstances also differ systematically between households with and without children.

A logistic regression analysis was conducted to estimate the extent to which these alternative explanations might account for the differential function of *Could not afford balanced meals* between households with and without children. Response to the item in question was regressed on a dummy variable for presence of children and a set of eight dummy variables representing raw score on the adult food security scale. Other household characteristics were then added to the model to see whether, and to what extent, the coefficient on the variable for presence of children was affected. The analysis was repeated with the sample limited to households in which all adults were nonelderly women.

Estimating the Effect of DIF on Prevalence Calculations

Three methods were used to assess the effect of DIF between households with and without children on estimated prevalence rates of food insecurity and very low food security. The first method used household measures based on item parameters estimated in separate models for households with and without children. Household parameters for households without children were based on item parameters adjusted to the metric of households with children, taking account of the lower discrimination parameter resulting from the adjustment. This assessed the joint effects of DIF of *Could not afford balanced meals* and the DIF (if any) of all other items, imposing only the constraint that the mean severity of the other items be the same in the two subpopulations. The second and third methods used household measures based on the confirmatory analysis in which item parameters for both subgroups were estimated in a single model, constraining all items except *Could not afford balanced meals* to function the same in the two subpopulations.

In the first two methods, bias on the prevalence of food insecurity was calculated as follows, and a similar process was followed for analysis at the very-low-food-security threshold.

1. A reference distribution across raw scores was calculated using data for households with no child, including extreme scores (0 and 10) and including all income levels. The observed (or measured) prevalence of food insecurity in this reference distribution was calculated as the percentage of the sample with raw scores 3 and higher.
2. An initial test threshold for food insecurity was selected about midway between the mean severity of raw scores 2 and 3.
3. Using the household parameters for households with children, the proportion of households in each raw score group with “true” severity higher than the test threshold (i.e., truly food insecure) was calculated based on a normal distribution with mean and standard deviation corresponding with those estimated for that raw score group. It was assumed that all households with raw score 0 were food

secure. Household parameters for households with raw score 10 were calculated as if the raw score were 9.5. (Assigning this conventional "pseudo-score" makes the mean and standard deviation of true values on the latent-trait estimable. Given the small percentage of households with this extreme score and its distance from the thresholds of interest, results are relatively insensitive to the exact value assigned.)

4. The proportions of truly food insecure in each raw score group were weighted by the proportion of the reference distribution with that raw score. The weighted sum across raw scores represented the "true" prevalence vis-à-vis the test threshold.

5. The test threshold was then iteratively adjusted (repeating steps 3 through 5) until the "true" prevalence was equal to the observed prevalence. This is the threshold at which there would be zero measurement bias (no difference between observed and true prevalence rates) for households with children if they had the same distribution across raw scores as the reference distribution.

6. Now, using the household parameters for households with no child, the proportion of households in each raw score group with "true" severity higher than the test threshold (i.e., truly food insecure) was calculated based on a normal distribution with mean and standard deviation corresponding with those estimated for that score group for households with no child. Again it was assumed that all households with raw score 0 were food secure, and household parameters for households with raw score 10 were calculated as if the raw score were 9.5.

7. The proportions of truly food insecure in each raw score group were weighted by the proportion of the reference distribution with that raw score. The weighted sum across raw scores represented the "true" prevalence of food insecurity in households with no child vis-à-vis the threshold at which measurement bias was zero for households with children.

8. The difference between the observed prevalence of food insecurity and the true prevalence of food insecurity in households with no child is an estimate of the bias of measured food insecurity in households with no child vis-à-vis households with children.

Within some range, the exact specification of a threshold for food insecurity is arbitrary. I initially selected a threshold to equate observed with "true" prevalence rates for one of the two groups. I then tested the robustness of the results across a range of thresholds to assess the sensitivity of results to the specification of the threshold.

The third approach used a somewhat simpler methodology. The proportion of households with raw score 2 on the other items (i.e., omitting *Could not afford balanced meals*) that would affirm the item, *Could not afford balanced meals*, was calculated for households with and without children, based on the respective severity parameters for that item in the two subpopulations and the estimated household severity parameter for raw score 2 on the "other items" scale. The difference in these proportions, multiplied by the proportion of households with raw score 2 on the "other items" scale, is an estimate of the bias due to DIF of *Could not afford balanced meals*. This analysis was repeated at raw score 5 to assess bias on very low food security.

Examining DIF Between Other Subpopulations

Item parameters were estimated in separate models using methods as described above for other subpopulations of interest. Several analyses were conducted based on household composition to assess both DIF due to household composition and DIF between men and women respondents (by comparing responses between households in which all adults were males and those in which all adults were females). Similarly, responses were compared among households with all adults in various age ranges. DIF assessments were conducted for:

- Multi-adult households with no child: Male-only versus female-only
- Women living alone versus no-child-multi-adult-female-only households
- Men living alone versus no-child-multi-adult-female-only households
- No-child households: All adults younger than 40 years versus all adults in the age-range 40-64
- No-child households: All adults ages 65 or older versus all adults in the age range 40-64
- No-child households: Northeast versus Midwest Census Region
- No-child households: South versus Midwest Census Region
- No-child households: West versus Midwest Census Region
- No-child households: Outside versus inside Metropolitan Statistical Areas
- Black-non-Hispanic versus White-non-Hispanic households (separate comparisons for households with and without children)
- Hispanic versus White-non-Hispanic households (separate comparisons for households with and without children)

In only one of these comparisons, Hispanic versus White-non-Hispanic households with children, did DIF of any items appear serious enough to be of practical concern. For that comparison, the effect of the DIF on prevalence estimates was calculated using methods similar to those described above for households with no child versus households with children.

Findings

DIF Between Households With and Without Children

Comparison of item-severity parameters between households with and without children suggested that at least one item, *Could not afford balanced meals*, differed in severity between the two groups (table 4-1 and fig. 4-1). The standard deviation of item parameters was also about 9 percent smaller in households with no child than in households with children when both models were estimated on a logistic metric (discrimination parameter 1.0). This indicates a slightly weaker fit of the data to the Rasch model for house-

Table 4-1

Item severity parameters estimated in separate models for households with children and households without children

Item	Household with child	Household with no child			
		Adjustment #1 (all items considered equivalent)		Adjustment #2 (all items except *Could not afford balanced meals* considered equivalent)	
	Severity parameter	Severity parameter	Difference from household with child	Severity parameter	Difference from household with child
Worried food would run out	3.17	3.67	0.50	3.46	0.29
Food bought did not last	4.36	4.48	.12	4.30	-.06
Could not afford balanced meals	5.34	4.69	-.65	4.52	-.82
Adult cut or skipped meals (1 or 2 vs. 0)[1]	6.68	6.61	-.07	6.51	-.17
Adult cut or skipped meals (2 vs. 0 or 1)[1]	7.21	7.05	-.16	6.97	-.24
Ate less than felt should	6.79	6.72	-.07	6.63	-.16
Hungry but did not eat	8.41	8.46	.05	8.43	.02
Lost weight	9.33	9.21	-.12	9.21	-.12
Adult did not eat for whole day (1 or 2 vs. 0)[1]	9.20	9.39	.18	9.40	.19
Adult did not eat for whole day (2 vs. 0 or 1)[1]	9.50	9.71	.22	9.74	.24
Mean—all items	7.00	7.00			
Std. deviation—all items	2.19	2.19			
Discrimination parameter based on adjustment #1[2]	1.00	0.91			
Mean—all items except *Could not afford balanced meals*	7.18			7.18	
Std. deviation— all items except *Could not afford balanced meals*	2.34			2.34	
Discrimination parameter based on adjustment #2[2]	1.00			0.87	
Number of cases	24,418	22,546		22,546	

[1]Tabled severity parameters for the two thresholds of the two trichotomous items are the "Rasch-Thurstone" or "50-percent probability" values. The value for the lower threshold is the estimated severity level at which the probability of a response in either of the two affirmative categories is .5. The value for the upper threshold (2 vs. 0 or 1) is the estimated severity level at which the probability of a response in the most frequent (3 or more months) is .5.

[2]All models were estimated on a logistic metric (discrimination parameter = 1). The tabled discrimination parameters for households with no child are those required to equate the standard deviation of items considered to be equivalent in households with and without children for purposes of that comparison.

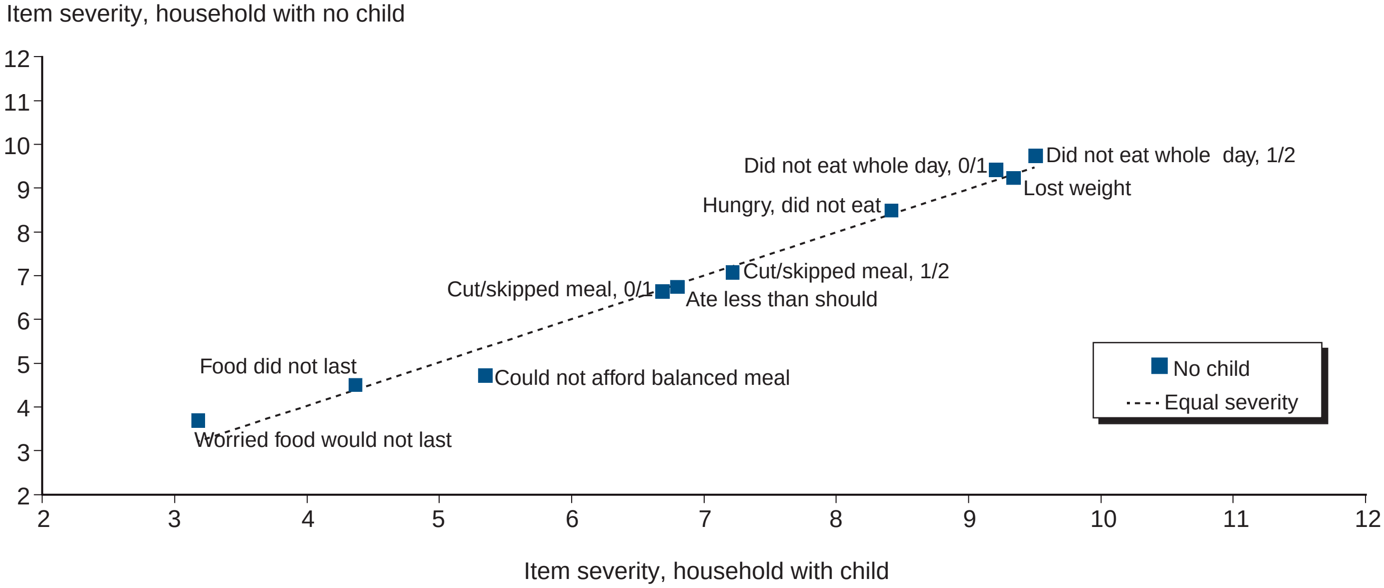
Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1997-2007 for households with incomes less than 185 percent of the Federal poverty line.

holds without children, that is, slightly lower average discrimination of items. Parameters for households with no children were adjusted for this difference prior to comparison (adjustment #1 in table 4-1). The signed area test also indicated that this item had the highest DIF, taking into consideration the distribution of households across raw score groups, and indicated possible, although somewhat weaker, DIF for the item *Worried food would run out* (table 4-2).

A second comparison was made in which the item *Could not afford balanced meals* was omitted from the set of items considered equivalent for purposes of adjusting item parameters (adjustment #2 in tables 4-1 and 4-2; fig. 4-2). With this adjustment, the severity parameter for *Could not afford balanced meals* was 0.82 logistic units lower for households with no child than for households with children. The standard deviation of parameters of this reduced set was 13 percent smaller for households with no child than for households with

Figure 4-1

Item severity comparison, households with no child versus households with child; estimates for households with no child constrained to equate mean and standard deviation of all items to those of households with child (adjustment #1 in table 4)



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007 for households with incomes less than 185 percent of the Federal poverty line.

Table 4-2

Signed-area test statistics for difference between households without children and households with children

Item	Adjustment #1 (all items considered equivalent)	Adjustment #2 (all items except *Could not afford balanced meals* considered equivalent)
	Mean difference in probability of "yes" response	
Worried food would run out	-.073	-.047
Food bought did not last	-.023	.003
Could not afford balanced meals	.102	.127
Adult cut or skipped meals (1 or 2 vs. 0)[1]	.011	.022
Adult cut or skipped meals (2 vs. 0 or 1)[1]	.023	.035
Ate less than felt should	.017	.029
Hungry but did not eat	.006	.008
Lost weight	.015	.015
Adult did not eat for whole day (1 or 2 vs. 0)[1]	-.004	-.005
Adult did not eat for whole day (2 vs. 0 or 1)[1]	-.004	-.005

Item severity parameters and scale discrimination parameters were those presented in table 4-1. Probabilities of affirmative responses were calculated for households with and without children based on respective item severity parameters and household severity parameters for each raw score. Household severity parameters were maximum likelihood estimates based on item severity parameters for households with children. Tabled values are weighted mean differences in probability of affirmative response across raw score groups, with weights based on the population distribution of households with children across raw scores. The mean difference corresponds to the signed area between the item characteristic curves.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007; item parameters were estimated from CPS-FSS households with annual incomes less than 185 percent of the Federal poverty line. Distribution of households with children across raw scores (used as weights in calculating the signed areas) are for households with children in all income ranges in the 2003-05 data.

children. Again, this difference was adjusted prior to the tabled comparison. Both the plot of item-severity parameters and the signed area test statistics indicated that any DIF in the remaining items was too small and dispersed among the items to identify any single item as its source.

Item infit statistics allay concerns about DIF in discrimination. Items could differ in discrimination as well as severity between households with and without children. I did not fit 2-parameter models to directly compare item discriminations, but item infit statistics provide a proxy (inversely) for discrimination and would identify any items with substantially different discrimination in the two groups. USDA is unlikely to use a 2-parameter model for food security measurement (as discussed in Chapter 3), but substantially different discrimination would still be a concern with regard to comparability. However, all item infit statistics were similar for households with and without children (table 4-3). With one exception, infit statistics were either both higher than 1 or both lower than 1 and of approximately equal magnitudes. The exception was the first item, for which both were very near unity.

Reasons for DIF Between Households With and Without Children

Only a relatively small proportion of the difference in severity of the item *Could not afford balanced meals* between households with and without children is due to characteristics of respondents or households, other than the presence of children, that were identified in the logistic regression analysis. With control only for the household's raw score on the scale, the logistic coefficient on the *household with child* dummy variable was -0.70 (corresponding to an odds ratio of 0.46;

Table 4-3

Item-infit statistics from separate models for households with children and households without children

Item	Households with child	Households with no child
Worried food would run out	0.99	1.01
Food bought did not last	.88	.89
Could not afford balanced meals	1.22	1.17
Adult cut or skipped meals (1 vs. 0)[1]	.93	.96
Adult cut or skipped meals (2 vs. 1)[1]	1.21	1.21
Adult cut or skipped meals (overall)[2]	1.17	1.22
Ate less than felt should	.74	.76
Hungry but did not eat	.79	.78
Lost weight	.85	.88
Adult did not eat for whole day (1 vs. 0)[1]	1.01	1.04
Adult did not eat for whole day (2 vs. 1)[1]	1.24	1.16
Adult did not eat for whole day (overall)[2]	1.25	1.20
Number of cases	24,418	22,546

[1]Fit statistics for trichotomous items for individual thresholds were calculated from mean square differences between observed responses and the expected mean response given that the response was in one of the two categories at that threshold.

[2]Overall fit statistics for trichotomous items were based on observed value (0, 1, or 2) versus expected value in the 0-2 range.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007 for households with incomes less than 185 percent of the Federal poverty line.

model 1 in table 4-4). The coefficient declined in magnitude to -0.58, a decline of 17 percent, with controls for proportion of adults who are female (a proxy for sex of respondent), income (quadratic), age of oldest adult (quadratic), race and Hispanic ethnicity of reference person, education of most highly educated adult, and residence relative to metropolitan statistical area. The decline was somewhat less, around 13 percent, when any effect due to the sex of the respondent was obviated by restricting the sample to households in which all adults were nonelderly females (models 3 and 4 in table 4-4). In both cases, the modest decline in the coefficient on *household with child* was accounted for almost entirely by inclusion of *age of the oldest adult* in the model.

DIF for this item was observed in all years from 1998 to 2007. Model 2 in table 4-4 was repeated with the addition of dummy variables for survey year and interactions between each survey year dummy and the *household with child* dummy variable (analysis not shown). The difference between households with and without children, calculated as the sum of the coefficients on the *household with child* dummy and the interaction variable for each year, ranged from -0.48 to -0.73 logistic units.

Table 4-4

Logistic regression models of response to item, *Could not afford to eat balanced meals* on household characteristics with controls for raw score on the Adult Food Security Scale

Characteristic[1]	All low-income households[2]		Low-income households in which all adults were nonelderly women	
	Model 1	Model 2	Model 3	Model 4
Intercept	-.77**	-1.71**	-.86**	-1.77**
Household with child	-.70**	-.58**	-.82**	-.71**
Proportion of adults who are female		-.11*		
Income/poverty ratio		.28*		.58*
(Income/poverty ratio)-squared		-.11		-.26*
Age of oldest adult		.021**		.013**
(Age of oldest adult)-squared[3]		-.00012*		
Black non-Hispanic household reference person		-.04		.05
Hispanic household reference person		.34**		.29**
Other/multiple race non-Hispanic		.14*		.13
Most highly educated adult less than high school		.09*		.04
Most highly educated adult some college, less than bachelor		-.05*		-.07
Most highly educated adult bachelor or other 4-year degree		-.03		-.03
Most highly educated adult graduate or professional degree		-.03		-.01
Resident in principal city of metropolitan statistical area		.01		-.04
Resident in metropolitan statistical area, not identified		.06		.13
Resident outside metropolitan statistical area		.03		.12
Number of cases	46,964	46,964	14,277	14,277

* Difference from zero statistically significant, p<.01

**Difference from zero statistically significant, p<.001

[1]All models included a set of 8 dummy variables, not shown, representing raw scores 2-9 on the aAdult Food Security Scale.

[2]Households with annual incomes less than 185 percent of the Federal poverty line.

[3]Age of oldest adult was modeled as linear in model 4 because the coefficient on the squared term was weak and not statistically significant (p=.52) and its inclusion made interpretation of the coefficient on the linear term difficult.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007 for households with incomes less than 185 percent of the Federal poverty line.

Effect of DIF on Prevalence Calculations

The lower severity of *Could not afford balanced meals* in households with no child was expected to bias upward prevalence estimates of food insecurity for households with no child relative to households with children. Classification as food insecure versus secure is based on a raw score of 3 or more. The item in question is the third item in severity order, and households with no child are more likely to affirm it than are households with children at the same level of true food insecurity.

The prevalence of very low food security was also expected to be biased upward somewhat for households with no child relative to households with children. Three items with severity parameters near the threshold for very low food security (6 or more affirmative responses) have severity parameters slightly lower for households with no child than for households with children (see fig. 4-2). Even though the extent of DIF for these items is slight, the combined effects of the three items plus the low severity of *Could not afford balanced meals* is likely to result in some degree of bias.

Analyses based on the mean and standard deviation of household measures confirm these expectations and provide estimates of the extent of biases. The measured prevalence of food insecurity in households with no child is estimated to be biased upward by 0.60 percentage points relative to households with children (table 4-5). The corresponding upward bias on very low food security is estimated to be 0.18 percentage points. These amount to about 7.2 and 5.2 percent, respectively, of the observed values.

The size of the estimated measurement bias between households with and without children is not sensitive to the selection of the threshold for "true" food insecurity across a reasonable range. If the threshold is set at a more severe level, the difference between measured and "true" prevalence increases for both types of households. However, the changes are of about the same magnitude, so the relative bias (i.e., between households with and without children) changes little if at all. Initial estimates were made with the threshold set at a level that resulted in zero bias for households with children. I re-estimated bias with the threshold set at a level that resulted in zero bias for households with no children. The estimated bias between the two types of households was essentially unchanged.

The estimated bias on the prevalence of food insecurity from the single-model analysis, which assessed the effect of DIF of only the single item, *Could not afford balanced meals,* was +0.43 percentage points, somewhat smaller than that described above. Bias on the prevalence of very low food security was considerably smaller, as expected, since the severity of *Could not afford balanced meals* is far below the threshold for very low food security, and the severity parameters of all other items were constrained to be equal in the two subpopulations.

Results from the third method, based on households with raw score 2 calculated from the remaining items after omitting *Could not afford balanced meals,* were similar to those just described (analysis not shown). Biases were +0.55 percentage points on the prevalence of food insecurity and +0.07 percentage points on the prevalence of very low food security.

Table 4-5

Bias on estimated prevalence of food insecurity and very low food security of households with no child versus households with children due to differential item function[1]

Method and characteristic	Measured pre-valence	True prevalence: Household with child	True prevalence: Household with no child	Bias, measured versus true: Household with child	Bias, measured versus true: Household with no child	Bias, No child versus with child: Percent of all house-holds	Bias, No child versus with child: Percent of observed pre-valence
	Percent	*Percent*	*Percent*	*Percentage points*	*Percentage points*	*Percent*	*Percent*
DIF of all items[2]							
Food insecurity	8.27	8.27	7.67	0.00	+0.60	+0.60	+7.21
Very low food security	3.45	3.45	3.27	0.00	+.18	+.18	+5.23
DIF of item *Could not afford to eat balanced meals* only[3]							
Food insecurity	8.27	8.27	7.84	0.00	+.43	+.43	+5.25
Very low food security	3.45	3.45	3.39	0.00	+.06	+.06	+1.96

DIF = differential item function.

[1]Calculations for both types of households (with and without children) were based on the observed distribution across raw score groups of all households without children (including all income levels). "True" prevalence rates are based on thresholds selected to equate true and measured prevalence for households with children. Thus the zero bias for households with no child is an artifact of this specific selection (arbitrary within a reasonable range) of the threshold. The bias of child versus no-child households is not sensitive to specification of this threshold within a reasonable range since the measured versus "true" bias of the two groups change by nearly equal increments with small changes in the threshold.

[2]Models were estimated separately for households with and without children. The mean and standard deviation of household severities for each raw score group were calculated from the estimated item parameters using maximum likelihood methods. The scale for households with no children was adjusted by a linear transformation to equate the mean and standard deviation of all items except *Could not afford to eat balanced meals* to those of the scale for household with children (adjustment #2 in table 4-1). The mean of household severities for each raw score group for households with no children were then adjusted by the same linear transformation, and the standard deviation was adjusted by the multiplicative constant.

[3]The model was estimated jointly for households with and without children. All items except *Could not afford to eat balanced meals* were constrained to have equal severity parameters in the two groups and item discrimination was constrained to be equal for all items in both groups.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007; scale characteristics were calculated using data for households with incomes less than 185 percent of the Federal poverty line. The reference distribution for estimating bias included households at all income levels.

Taking account of the bias due to DIF does not change the qualitative findings of national comparisons. The national prevalence of food insecurity (based on the adult scale) was 12.9 percent for households with children and 8.2 percent for households with no child. The difference of 4.7 percent would be larger by 0.6 percentage points if corrected for the estimated bias due to DIF. The national prevalence of very low food security was 4.0 percent for households with children and 3.3 percent for households with no children. Based on the calculated bias due to DIF, the difference of 0.7 percentage points may understate the true difference by 0.18 percentage points.

DIF Between Other Subpopulations

Plots of item severity between the various subpopulations analyzed are included in appendix C. With one exception, the relative severity of items is very consistent among these subpopulations. Infit statistics (not shown) were also generally similar.

The single exception was a modest degree of differential function of several items between Hispanic households with children and White non-Hispanic households with children (fig. C-13). This may be due to linguistic and cultural differences, although similar differences were not found for households with no children. Again, *Could not afford balanced meals* was the item that appeared to have the largest DIF, but *Cut or skipped meal* (especially at the 0/1 threshold) and *Ate less than should* also appeared to be large enough to be of possible concern. I used the first method described previously for households with and without children to estimate the extent of bias on prevalence estimates. In spite of the lower severity of *Could not afford balanced meals*, the bias on estimated prevalence of food insecurity in households with children was slightly downward (by 0.14 percentage points) for Hispanic versus White non-Hispanic households (analysis not shown). This is apparently due to the higher severity of *Cut or skipped meal* and *Ate less than should* and to the higher overall discrimination of items in Hispanic households. The downward bias on estimated prevalence of very low food security was somewhat larger, amounting to 0.72 percentage points (14 percent of the "true" value).

Adjusting for the bias due to DIF does not change the qualitative findings of higher food insecurity among Hispanic households with children compared with White non-Hispanic households with children. The national level prevalence rates (based on the adult scale) were 20.7 percent for Hispanics and 9.3 percent for White non-Hispanics. The additional difference that would result from correcting for estimated bias due to DIF would be 0.14 percentage points. Bias due to DIF was somewhat more substantial relative to estimated percentages for very low food security. Estimated prevalence rates were 5.2 percent for Hispanic households and 3.2 percent for White non-Hispanic households. The true difference may be greater by 0.7 percentage points.

Conclusions

The key measurement characteristics of items in the Adult Food Security Scale are remarkably similar across subpopulations that differ by age, sex, race, household composition, Hispanic ethnicity, and region. In general, prevalence rates of food insecurity and very low food security can be compared across subpopulations assessed in this study with relatively little concern about bias.

DIF does exist between households with and without children, and the difference is somewhat greater for Hispanic households. DIF is most pronounced for the item *Could not afford balanced meals*, and other items appear to be slightly affected. Only a small part of the DIF for *Could not afford balanced meals* is explained by differences in the sex, age, education, income, race, Hispanic ethnicity, or rural-urban residence of respondents. It seems likely, therefore, that it results from differences in strategies for managing food insecurity in the two types of households rather than from differences in understanding of the objective conditions referred to by the question.

The practical implications of the DIF as estimated in this study are modest, and correcting for them would not change the assessment of intergroup differences in food security in any important ways. The current classification method (that does not take account of DIF) is estimated to overstate the prevalence of food insecurity in households with children relative to households with no child by about 0.6 percentage points. The corresponding bias on very low food security is estimated to be 0.18 percentage points. These biases are relatively small compared with prevalence estimates, and correcting for them would increase the already substantial intergroup differences as estimated by current methods. The bias on estimated prevalence of food insecurity is also small relative to the bias because of the nonequivalent thresholds on the 18-item and 10-item scale, and partially offsets that bias.

The results of this study could be used in various ways by USDA depending on decisions it makes on other aspects of the measurement methods. National-level estimates could be adjusted for the estimated bias due to DIF, but this might not be advisable. Even though the statistical basis for these adjustments may be quite strong, making such adjustments might raise questions about the validity of the comparisons, introducing something of a “black box” quality at the expense of the transparency and ease of explanation of the current method. It may be more appropriate to use the estimates from this study to comment on observed relationships, or to give an approximation of what bias between two types of households may be, in cases where the bias appears to be large enough to affect the conclusions of a study or public perceptions of the importance of a reported difference in prevalence rates.

Chapter 5

Assigning Food Security Status Probabilistically To Account For Measurement Error

Abstract

This chapter assesses the recommendation by the Committee on National Statistics (CNSTAT) that USDA "… develop a new classification system that reflects the measurement error inherent in latent variable models. This can be accomplished by classifying households probabilistically along the latent scale, as opposed to the current practice of deterministically using the observed number of affirmations." Prevalence rates calculated using the current standard methods (which assign food security deterministically, or discretely) are compared with rates calculated using the more precise probabilistic methodology recommended by CNSTAT. With one exception, trends over time and comparisons across key population subgroups are essentially undistorted by the current methodology. The exception is that the higher prevalence of food insecurity for households with children relative to households without children is exaggerated by the current methodology, although both methods show substantially higher prevalence among households with children. Use of the probabilistic methodology would reduce measurement error on prevalence estimates substantially, particularly for food insecurity among children.

Background

The Committee on National Statistics (CNSTAT) panel that reviewed USDA's food security measurement methods recommended that USDA "…develop a new classification system that reflects the measurement error inherent in latent variable models. This can be accomplished by classifying households probabilistically along the latent scale, as opposed to the current practice of deterministically using the observed number of affirmations. Furthermore, the new classification system should be more closely tied to the content and location of food insecurity items along the latent scale." (NRC, 2006). Recognizing that this more complex method might not be practically accessible to researchers in State and local surveys, the panel also stated, "USDA should study the differences between the current classification system and the new system, possibly leading to a simple approximation to the new classification system for use in surveys and field studies." (NRC, 2006).

The current standard U.S. measure of household food security assigns each surveyed household to a specific single food security status (high, marginal, low, or very low) depending on the number of food-insecure indications they report—that is, on their raw score on the set of food security items. For example, households with raw score 2 are assigned to marginal food security, while those with raw score 3 are assigned to low food security. The range for very low food security is raw score 6 or higher for households without children present and raw score 8 or higher (considering both adult and child questions) for households with children present. (For most reporting purposes, the high and marginal food security categories are combined and reported as food secure. For some purposes, the low and very low food security categories are combined and reported as food insecure.)

Latent-trait measurement is, however, inherently probabilistic. Households in each raw score group do not all have the same "true" level of severity of food insecurity, but rather a range of true levels of food insecurity clustered around the maximum-likelihood-estimated level on the latent trait associated with that raw score. The measurement model supports calculation of the mean and standard deviation of true food insecurity of households within each raw score. The CNSTAT recommendation is that USDA use this measurement error information in its calculation of food security prevalence rates.

In principle, probabilistic assignment of food security status provides a more precise measure of food security than does discrete assignment since the former utilizes more information (e.g., 19 raw score groups rather than 3 or 4 status categories). For example, both raw score 3 and 4 are classified as having low food security, but the probability that a household with raw score 4 is correctly so classified is higher than the corresponding probability for a household with raw score 3. Comparisons over time or across subpopulations could, in principle, be distorted by discrete assignment if the distribution of households across raw score groups changes over time or differs across subpopulations. Probabilistic assignment also allows more precise comparison of food security measurement between households with children present (based on 18 items) and households with no child present (based on 10 items). Comparability based on deterministic assignment relies on breaks between raw score groups occurring at the same level of severity, which may not always be realized (and clearly is not realized in the case of the food-insecure threshold on the household food security scale).

The central question for USDA (and the Federal food security measurement project) is whether the advantages of probabilistic assignment outweigh the loss of transparency and ease of explanation of the discrete assignment method. Using a probabilistic assignment methodology would create more of a "black box" impression of the measure. Researchers may be comfortable with latent-trait measurement, but popular and policy audiences may not find the concepts intuitively accessible. An important factor in assessing this tradeoff is the extent to which probabilistic assignment would improve the measure. Put the other way round, the key questions are "To what extent are our present impressions of trends over time and comparisons across economic and demographic subpopulations distorted by reliance on the simpler discrete assignment?" and "Are the distortions serious enough to justify use of a more complex, less transparent methodology?"

Even if USDA decides to continue basing prevalence estimates on discrete assignment, the probabilistic methods could be used to assess the statistical significance of differences in prevalence rates between subpopulations and changes in prevalence rates over time. Thus, a secondary question is, "To what extent is measurement error reduced by use of probabilistic rather than discrete assignment of food security status?"

To answer these questions, prevalence rates of food insecurity and very low food security over time and across selected subpopulations were compared, based on discrete (current method) and probabilistic (more precise method) assignment of food security status. These analyses were conducted for three scales: the 18-item household scale (which includes adult and child items), the 10-item adult scale, and the 8-item child scale, thus examining effects on all of the prevalence rates commonly reported by USDA. If prevalence rates based on the two methods differ little, then the discrete method is preferred for its transparency and simplicity. However, if reliance on that method is distorting our understanding of how food security has changed over time or our understanding of differences in prevalence across important economic and demographic subpopulations, then the probabilistic method may be justified in spite of its greater complexity.

The extent of measurement error between the two methods for each of the three scales was then compared. If measurement error is reduced substantially by using probabilistic assignment, then this method may be justified for assessing the statistical significance of differences and changes, since there is little cost in either effort or communicability in doing so.[1]

[1]The probabilistic methodology is explored in combination with two of the other potential enhancements recommended by the CNSTAT panel in Chapters 3 and 6.

Data and Methods

Data

The analyses used annual nationally representative survey data from the Current Population Survey Food Security Supplements (CPS-FSS) for the years 1998-2006 and 2010. Households that received nonstandard questions (questions under test) were omitted from the sample. This included households in month-in-sample 8 of the 1998 and 1999 surveys if they had more than one child or more than one adult and all households in month-in-sample 8 of the 2000 survey.[2] Households with no valid responses to any of the food security scale questions (about 0.3 percent of the sample) were omitted from the entire analysis. The sample for the main analyses, comparing the two food security measurement methods over time and across subpopulation, consisted of 398,901 households.

[2]Households in month-in-sample 8 in the 1998 and 1999 surveys with more than one child were administered nonstandard food security questions referenced to a specific (randomly selected) child rather than "any of the children." Households with more than one adult were administered nonstandard food security questions referenced to the respondent rather than to "you or other adults in the household." Households in month-in-sample 8 of the 2000 survey were administered nonstandard versions of the initial three food security questions that were being tested for Health Canada.

The sample used for estimation of item and household parameters was further limited to those with complete responses (less than 0.6 percent of those with any valid response had a missing response to any item) and with nonextreme responses. The omission of extreme responses reduced the sample sizes considerably, since a large majority of U.S. households are fully food secure. Sample sizes for item and household parameter estimation were as follows: 68,564 for the 10-item adult scale; 35,582 for the 18-item household scale (which includes adult and child items); and 21,866 for the 8-item child scale.

Data from the December 2010 CPS-FSS were used to assess the extent to which measurement error in prevalence rates would be reduced by using probabilistic assignment of food security status rather than the standard discrete assignment now used by USDA. The 2010 data were used for this analysis because 2010 was the first year for which the Census Bureau calculated replicate weights for the Food Security Supplement to support calculation of sampling errors using balanced repeated replication (BRR) methods.

Household sampling weights for the food security supplement were used for all calculations.

Food Security Status—Discrete Assignment

Assignment of food security status discretely (i.e., each household assigned completely to a specific status based on raw score) followed the standard methods (Bickel et al., 2000; Nord and Bickel, 2002):

- Household scale
    - Items—This is often described as an 18-item scale. More precisely, it consists of 15 items, of which 3 are trichotomous and 12 dichotomous. For households with no child present it consists of the eight adult- and household-referenced items, of which two are trichotomous and six dichotomous.
    - Coding—Households with raw score 3 or higher were classified as food insecure. Very low food security was identified by raw scores of eight or higher for households with children present and six or higher for households with no child present.
- Adult scale
    - Items—This is often described as a 10-item scale. More precisely, it consists of eight items, of which two are trichotomous and six dichotomous. It is the same scale described as the household scale for households without children.
    - Coding—Households with raw score 3 or higher were classified as food insecure. Households with raw score 6 or higher were classified as having very low food security.
- Child scale
    - Items—This is often described as an 8-item scale. More precisely, it consists of the seven child-referenced items, of which one is trichotomous and six dichotomous.
    - Coding—Households with raw score 2 or higher were classified as having food insecurity among children. Households with raw score 5 or higher were classified as having very low food security among children.

Food Security Status—Probabilistic Assignment

Each household with a nonextreme raw score on a given scale (i.e., omitting raw scores zero and maximum for that scale) was assigned a probability of being food insecure and a probability of having very low food security on that scale. Households with raw score zero were assigned probabilities of zero, and those with maximum raw scores were assigned to the very low food security category with certainty.[3] The estimated prevalence of food insecurity or very low food security of the population or subpopulation was then calculated as the mean of the respective probabilities in the sample or subsample.

The probability that a household was food insecure was calculated using the following procedure. (The example is described for food insecurity on the adult scale. Similar procedures were followed for very low food security and for the household and child scales.)

[3]The CML Rasch-model estimation methods used for this chapter cannot provide parameter estimates or measurement errors for extreme raw scores. Estimates using other methods suggest that a small proportion of households with raw score zero may be food insecure, but the proportion cannot be estimated with any confidence since information about the distribution of the latent trait among the large share of households with raw score zero is not available. At the other end of the scale, only a tiny proportion of households with maximum raw scores are misclassified by assuming that all have very low food security. The proportion of all households with such high raw scores is also very small, so assigning them with certainty to very low food secure status introduces only negligible distortions.

1. Item parameters were estimated using conditional maximum likelihood (CML) estimation methods that took appropriate account of the trichotomous items and of the screening within the food security module.[4] The estimation sample included all households with complete, nonextreme responses in all years 1998-2006. (Item parameters are provided in appendix table D-1.)

2. Household severity levels (i.e., location on the latent trait) and measurement error were calculated for each raw score using maximum likelihood methods. (Household parameters and measurement errors are provided in appendix table D-2.)

3. An initial threshold for the insecure range was approximated, and the number of households in each raw score that would be expected to have true levels of food insecurity in excess of the threshold were calculated assuming a normal distribution with mean at the household parameter estimate for that raw score and a standard deviation equal to the measurement error for that raw score. (The full-sample distributions of households across raw score groups for each scale are provided in appendix tables D-2, D-3, D-4, and D-5.)

4. The number of households in the food-insecure range were totaled across raw score groups. The threshold for the food-insecure range was then adjusted iteratively until the number of households in the food-insecure range was equal to the number with raw score 3 or higher. (The probabilities of food insecurity and very low food security for each raw score group are provided in appendix tables D-2, D-3, D-4, and D-5.)

5. Finally, the probability of food insecurity for each raw score was calculated based on the household parameter and measurement error for that raw score and the final threshold level for food insecurity. These probabilities were attached to each household record in the analysis data set.

The procedure for calculating probabilities of food insecurity on the household scale was somewhat more complex because different scales are used for households with and without children. The number of households that would be expected to have true levels of severity within the food-insecure range were calculated separately for each raw score group of households with children and for each raw score group of households without children. Calculation within each group was based on the number of households observed in that specific raw score group and the household parameter and measurement error for that specific raw score group. However, the same threshold for food insecurity was used both for households with and without children, and it was adjusted iteratively to get the same combined total of food-insecure households, but not necessarily the same number of food insecure households with children or without children. Specification of the same threshold for households with and without children was essential because an important question is whether the discrete assignment method biases comparisons between households with and without children.

[4]To reduce respondent burden in the CPS, and to avoid annoying respondents with inappropriate questions, respondents who say “never” to the first three questions (or five questions in households with children) are not administered the remaining food security questions, and are assumed to have high food security. A similar screen is assessed prior to asking about not eating for a whole day. Respondents who have said “no” or “never” to the questions since the first screen are not administered the remaining items, and responses of “no” are assumed in assigning food security. Ignoring these screens has minimal effects on item parameters, but distorts item-infit statistics slightly and item-outfit statistics greatly.

Comparing the Measurement Error of Prevalence Estimates Based on Discrete and Probabilistic Assignment of Food Security Status

Measurement errors of estimated prevalence rates of food insecurity and very low food security based on discrete and probabilistic assignment of food security status were calculated using two replication methods. Jackknife replication methods were used for the 2004-06 data and balanced repeated replication (BRR) methods for the 2010 data. Both of these methods are designed for the primary purpose of estimating sampling error, but in fact, both methods measure the combined sampling error and measurement error. Indeed, the two sources of error cannot be differentiated by these methods. In each analysis in this study, sampling error was the same for the two measurement methods, since they were based on the same data, so differences in Jackknife- or BRR-estimated errors can be ascribed to differences in measurement error.

The jackknife estimates used the eight month-in-sample "rotation" groups in each year, since these are, effectively, independent samples of U.S. households. Standard jackknife replication methods were applied to these subsamples to calculate variance estimates for prevalence rates of food insecurity and very low food security for each of the three scales for each year from 2004 to 2006. Averages across the 3 years were then calculated and compared.

The BRR methods were implemented in the SAS SurveyMeans procedure using replicate weights provided by the Census Bureau for the December 2010 CPS-FSS.[5] National-level BRR estimates were calculated as well as estimates for low-income households (those with incomes less than 185 percent of the Federal poverty line) and for low-income subpopulations by race and Hispanic ethnicity, household composition, and metropolitan-nonmetropolitan residence.

[5]Replicate weights were first provided for the CPS-FSS in 2010. The weights are based on the Fay method and, following specifications provided by the Census Bureau, the option "Fay=.5" was specified in the SAS SurveyMeans procedure.

Comparing Prevalence Rates Based on Discrete and Probabilistic Methods Over Time and Across Subpopulations

The procedures by which probabilistic-assignment thresholds were calculated produced prevalence rates that were identical to those based on the discrete assignment method for the combined sample including all years and all household types. Comparisons of discrete- and probabilistic-based prevalence rates over time and across population subgroups, then, addressed the question of whether, and to what extent, current discrete methods distort our understanding either of time trends in food insecurity or of the relative vulnerability of various subpopulations to food insecurity.

The subpopulations analyzed are those for which food security prevalence statistics are provided in USDA's annual report series, *Household Food Security in the United States* (see for example, Coleman-Jensen et al., 2011). The variables were calculated from data elements collected in the core labor force section of the Current Population Survey using the same procedures as those used in the annual reports.

Relating Categorization Procedures to the Severity of Specific Items

Although it was not the main purpose of this study, the relationships of current thresholds for food insecurity and very low food security were examined with the severity parameters of specific items in the scales. This is a first step toward addressing the CNSTAT recommendation that, "… the new classification system should be more closely tied to the content and location of food insecurity items along the latent scale." The team that did the original developmental work on the food security scale described in some detail the relationship of the conceptual categories of food insecurity to the cognitive content of scale items (Hamilton et al., 1997a; Hamilton et al., 1997b). Comparing the probabilistic thresholds implied by the current discrete assignment methodology with the severity parameters of the items that were considered to be "threshold" items for a given range of severity assesses how well those concepts are operationalized in the discrete methods currently used. Comparing prevalence rates based on thresholds set at the severity of those items with rates based on the current methodology assesses the substantive importance of the differences.

This part of the analysis is not meant to imply that the severity of the identified threshold items are the "right" severity levels for the thresholds. They provide one method for relating the current thresholds to the cognitive content of specific items. The approach taken by the team that originally created the food security measure related the thresholds to the cognitive content of the items through a different mechanism, based on the minimum conditions that would have to be reported for a household to be classified in a specific range of severity. So, for example, households with raw score 2 were characterized as those that affirmed the two least severe items and denied the rest. Households with raw score 3 were characterized as those that affirmed the three least severe items and denied the rest. It was the opinion of the experts that the latter should be classified as food insecure, but not the former.

Findings

Trends From 1998 to 2006

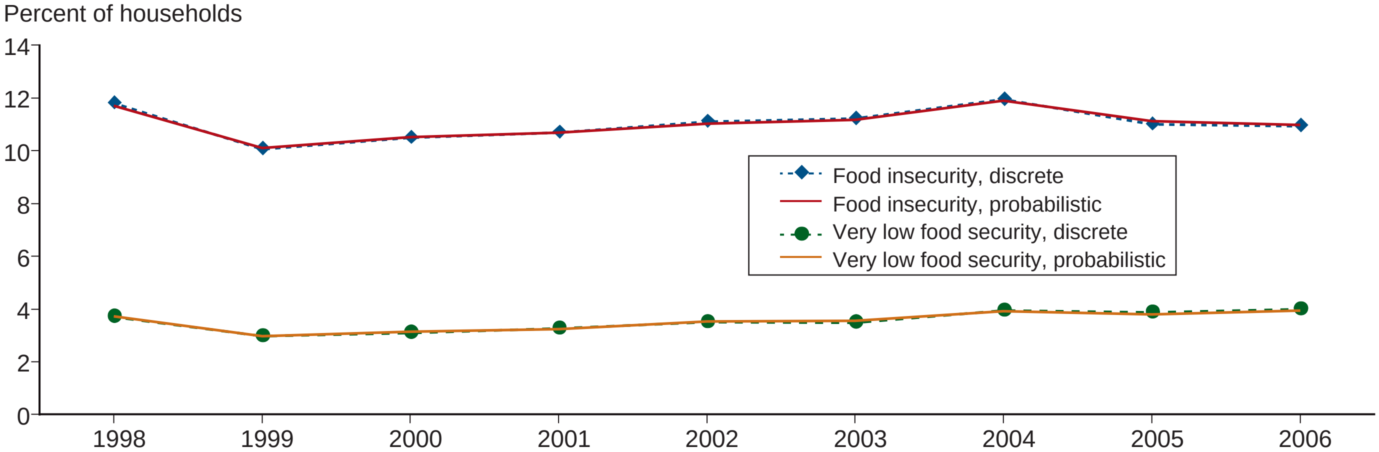
Trends in the prevalence of food insecurity and very low food security from 1998 to 2006 are virtually identical whether estimated by discrete or probabilistic assignment of households' food security status (figs. 5-1 to 5-3). There were some small differences in a few years, particularly noticeable for children's food insecurity in 2000. If the probabilistic assignment is taken as the standard, then the deterministic methods currently used certainly do not distort our understanding of year-to-year changes in food insecurity in any important ways.

Household Composition

The household scale (including household, adult, and child items) overstates the prevalence of food insecurity in all categories of households with children relative to all those without children (fig. 5-4). Based on the discrete assignment method, the prevalence of food insecurity in households with children was about double that in households without children. The probabilistic method indicates that the difference was only 65 percent.

Figure 5-1

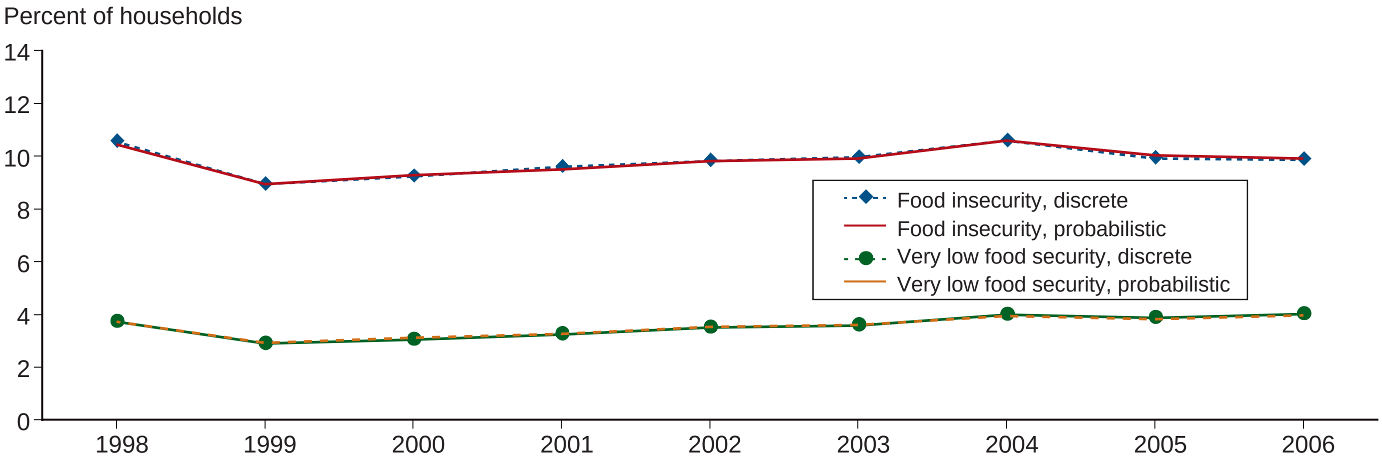
Prevalence of food insecurity and very low food security based on discrete and probabilistic assignment by raw score on the 18-item household scale, 1998-2006



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2006.

Figure 5-2

Prevalence of food insecurity and very low food security based on discrete and probabilistic assignment by raw score on the 10-item adult scale, 1998-2006

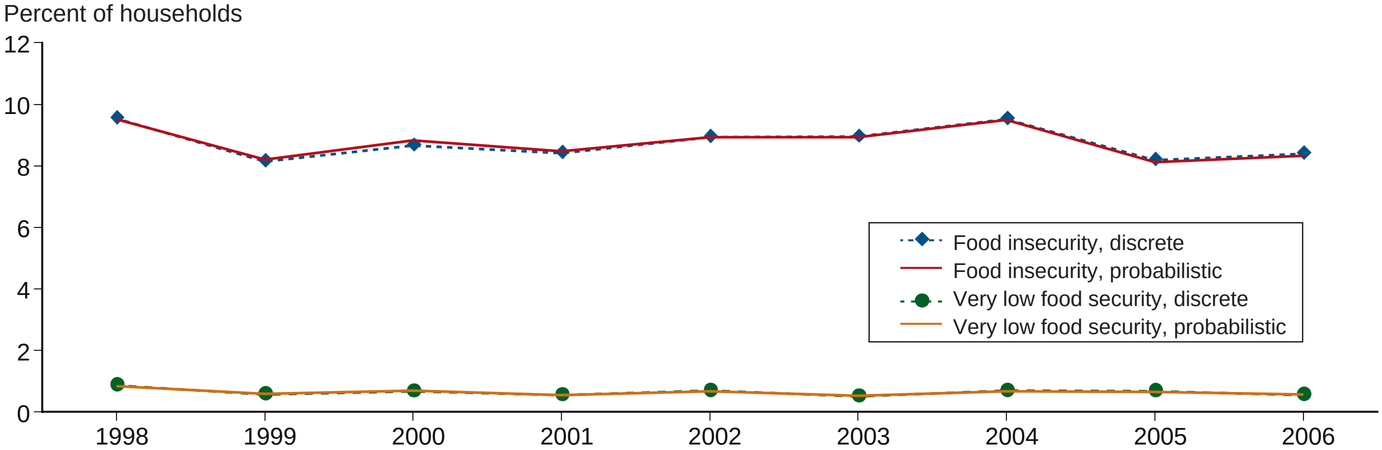


Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2006.

Figure 5-4 gives the impression that food insecurity is overstated in households with children and understated in households without children. There is, however, no objective standard for this determination. The prevalence rates based on probabilistic assignment in figure 5-4 result from the specific threshold that was selected for probabilistic assignment. That threshold was selected so as to result in a total population prevalence equal to that measured by the discrete assignment. The threshold is, in effect, a weighted average of the two thresholds implied by the two different discrete assignment procedures. All that can be said with statistical rigor is that the current classification method overstates the prevalence of food insecurity in households with children *relative to those without children* or, conversely, that it understates the prevalence of food insecurity in households without children *relative to those with children*.

Figure 5-3

Prevalence of food insecurity and very low food security based on discrete and probabilistic assignment by raw score on the 8-item child scale, 1998-2006



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2006.

This is a well-understood weakness of the current methodology, and it results from the fact that no discrete raw-score threshold is available on the 18-item scale applied to households with children that is precisely equivalent to that of raw score 3 on the 10-item scale applied to households without children. (See, for example, the discussion of the issue in footnote 12 of Coleman-Jensen et al., 2011.) Raw score 3 on the 18-item scale is not stringent enough, while raw score 4 is too stringent. If USDA continues to use the combined adult-child scale as a primary monitoring tool, this measurement artifact would be the single strongest argument in favor of using probabilistic rather than discrete assignment. No corresponding distortion is evident for very low food security. The two different thresholds (6+ for households without children and 8+ for households without children) are essentially at the same levels of severity and so introduce no distortions in the discrete assignment-based estimates.

The differences in prevalence rates between discrete and probabilistic assignment were large enough to affect relative rankings of some household types, although no substantive reversals resulted. The prevalence of food insecurity was higher for married couples with children than for households with no children based on the discrete method, but very nearly the same for the two groups, based on probabilistic assignment. The prevalence of food insecurity was very nearly the same for married couples with children and for women living alone and men living alone, based on deterministic assignment, while the prevalence was lower for married couples with children than for women or men living alone, based on probabilistic assignment.

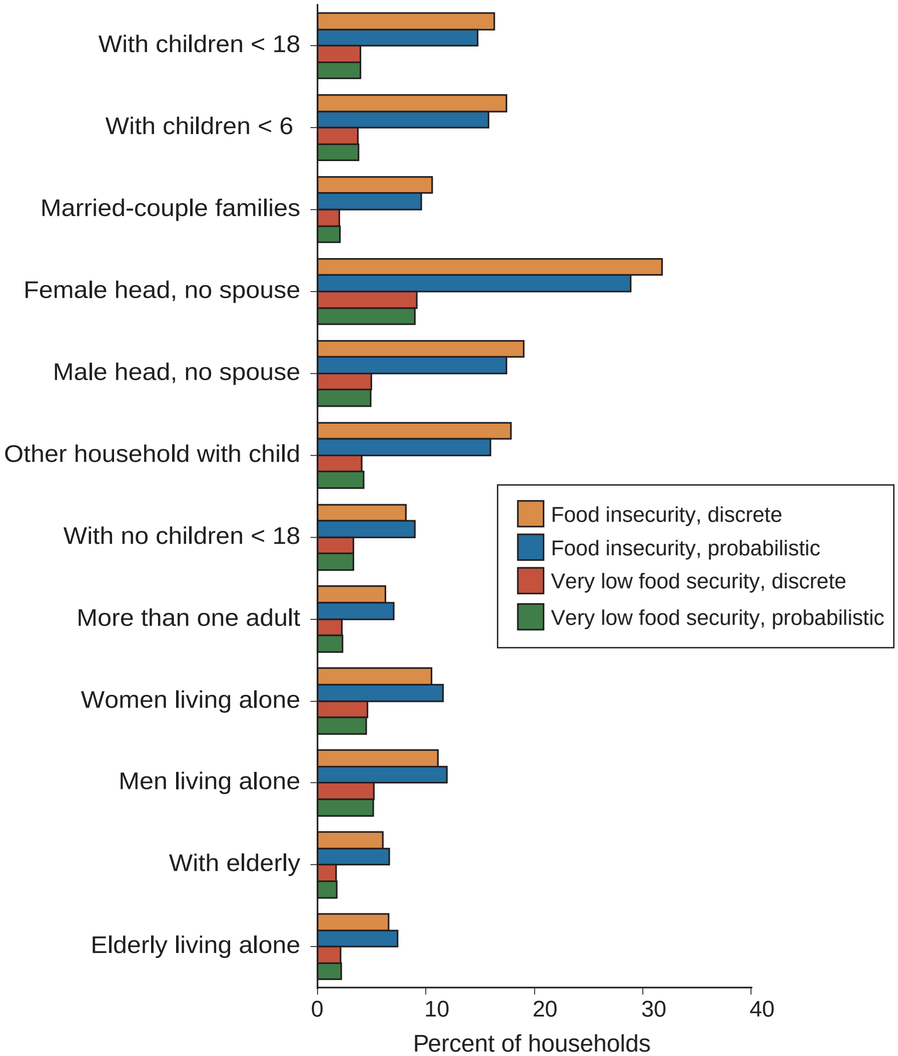
For the adult and child scales, differences were negligible between prevalence rates based on discrete and probabilistic assignment of food security status (figs. 5-5 and 5-6).

Race and Hispanic Ethnicity

Prevalence rates of food insecurity and very low food security across race and Hispanic ethnicity groups differed little between discrete and probabilistic

Figure 5-4

Prevalence of food insecurity and very low food security based on discrete and probabilistic assignment by raw score on the 18-item household scale, by household composition, average 1998-2006



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2006.

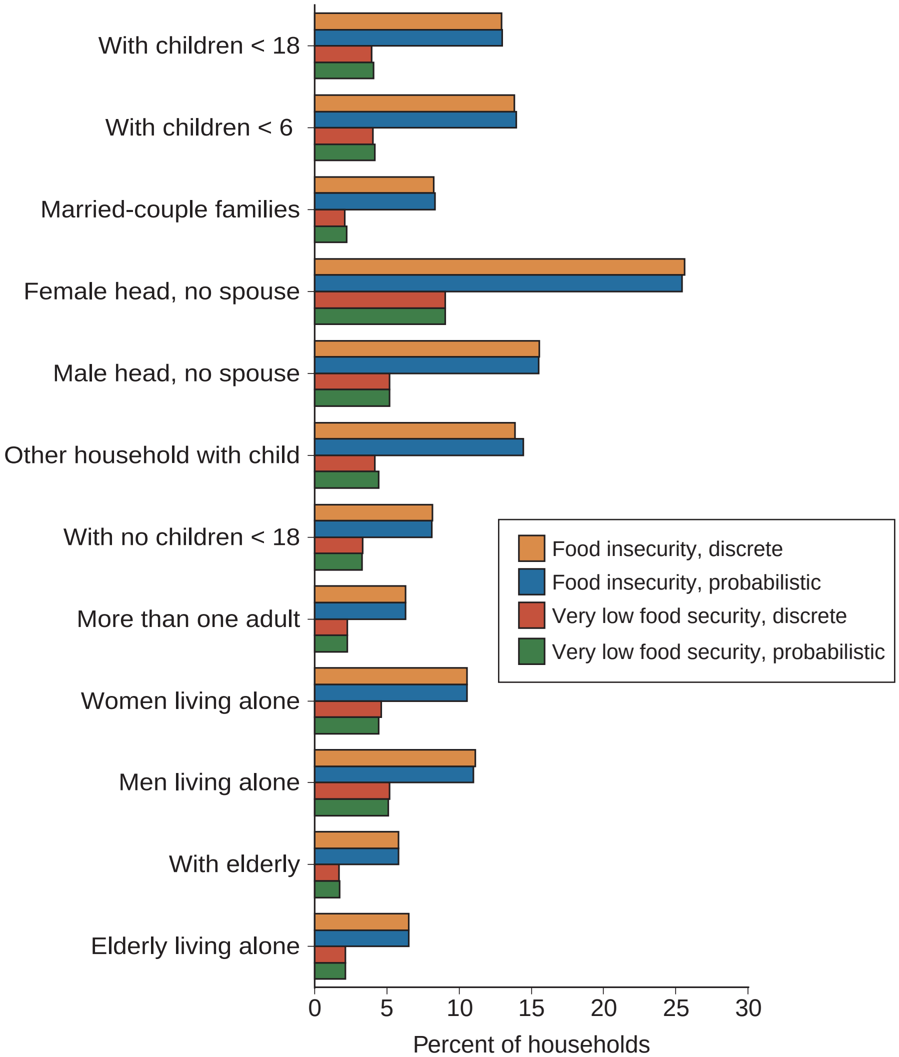
assignment methods (figs. 5-7, 5-8, 5-9). Discrete methods overstated the prevalence of food insecurity slightly for Black and Hispanic households relative to others, but the differences were very small compared with the differences between either of those groups and White non-Hispanics or other race-ethnic groups. The differences that did exist reflected primarily the larger proportions of households with children among Black and Hispanic households.

Income

Discrete assignment of food insecurity status slightly overstated the difference in prevalence rates between low-income and high-income households on all three scales (see figs. 5-7, 5-8, 5-9). However, the differences between methods were quite small relative to the difference between income groups. For example, based on the household scale, the difference between the lowest

Figure 5-5

Prevalence of food insecurity and very low food security based on discrete and probabilistic assignment by raw score on the 10-item adult scale, by household composition, average 1998-2006



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2006.

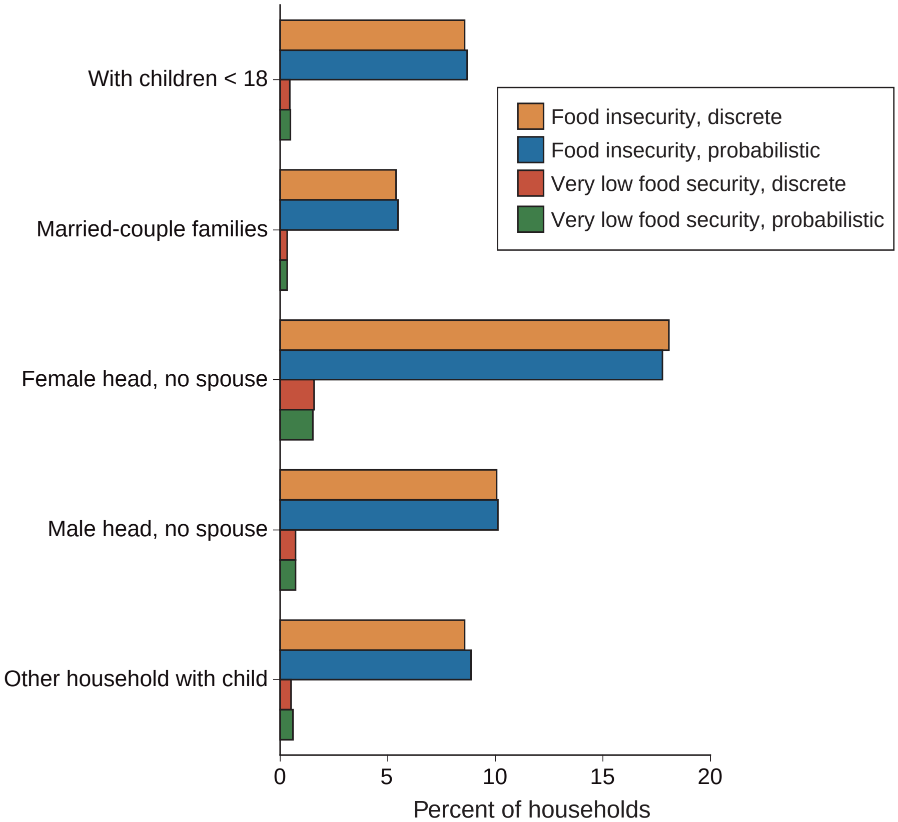
income (below poverty) and highest income (above 185 percent of poverty) categories typically reported by USDA is 31.7 percentage points based on discrete assignment and 30.5 percentage points based on probabilistic assignment. The discrete method, therefore, overstated the difference by 1.2 percentage points while the "true" difference was 30.5 percentage points. The differences between the two methods across income groups reflected primarily the larger proportions of households with children among the lower income groups.

Metropolitan Residence and Census Region

Comparisons of prevalence rates of food insecurity and very low food security across residence areas defined relative to Metropolitan Statistical Areas

Figure 5-6

Prevalence of food insecurity and very low food security among children based on discrete and probabilistic assignment by raw score on the 8-item child scale, by household composition, average 1998-2006



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2006.

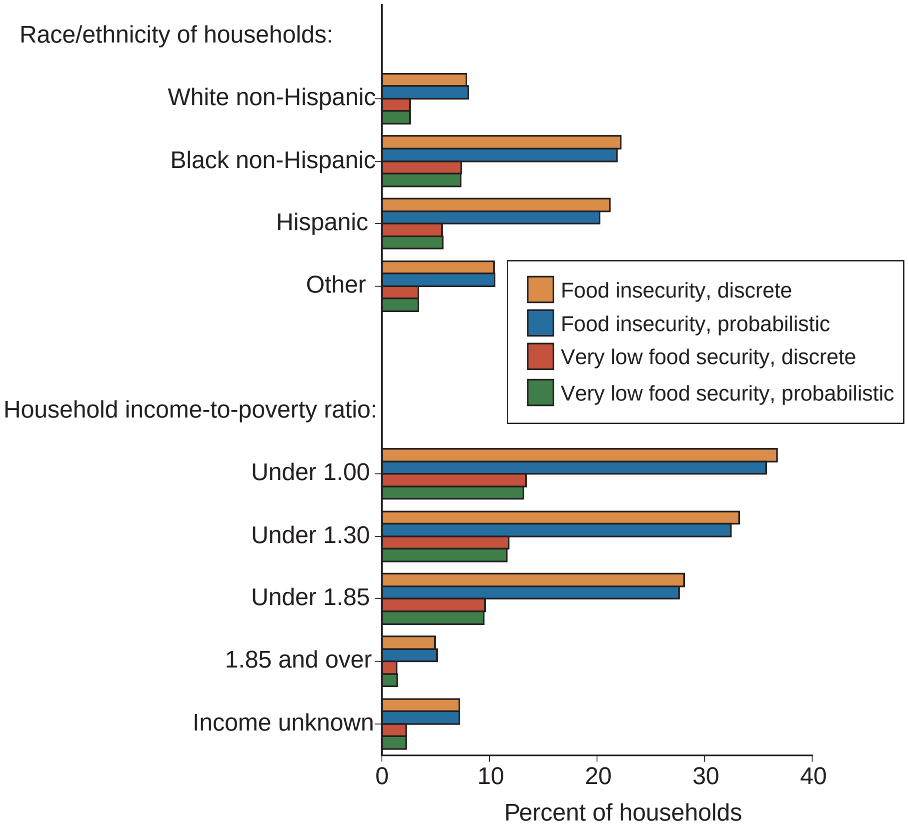
and Census regions were essentially unaffected by discrete versus probabilistic assignment of food security status (figs. 5-10, 5-11, 5-12).

Measurement Error

As expected, measurement errors based on probabilistic assignment of food security status were somewhat smaller than those based on discrete assignment. At the national level, the differences amounted to 5 to 8 percent for household and adult food insecurity and 12 to 15 percent for household and adult very low food security (table 5-1). The differences in measurement error for prevalence rates of children's food insecurity and very low food security were greater—16 to 17 percent for food insecurity and 31 to 33 percent for very low food security.

Differences in measurement errors for low-income households (those with incomes less than 185 percent of the Federal poverty line), were similar to those for all households. Measurement error for adult food insecurity was 10 percent smaller, and that for adult very low food security was 14 percent smaller, based on probabilistic assignment compared with discrete assignment (analysis not shown). Across low-income subpopulations, almost all reductions were in the 10- to 20-percent range.

Figure 5-7
Prevalence of food insecurity and very low food security based on discrete and probabilistic assignment by raw score on the 18-item household scale, by race, Hispanic ethnicity, and income, average 1998-2006



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2006.

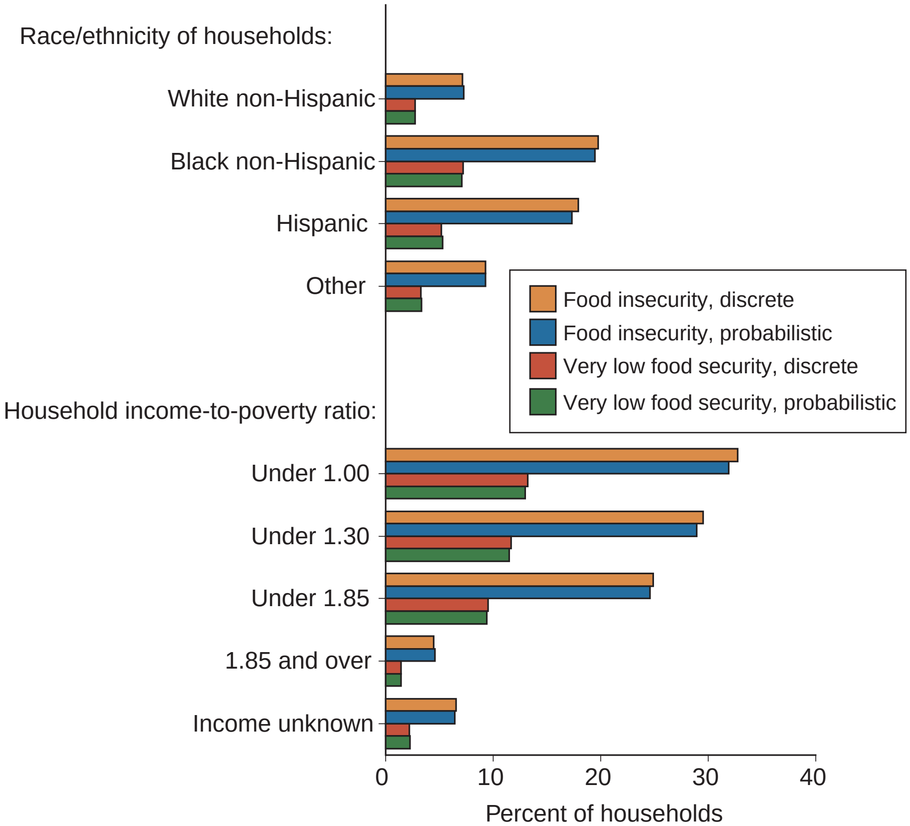
Variances as estimated by these replication methods include both sampling and measurement error. In these analyses, estimates based on the two measurement methods are subject to the same sampling error since each pair of estimates is based on the same sample, but measurement error is smaller for the probabilistic assignment method, which uses more of the available information in the response data. The percentage differences described above represent the difference in measurement error *as a percentage of combined measurement and sampling error.* The percentage reduction in measurement error, if it could be disentangled from sampling error, would be considerably larger.

Assessing the Location of Thresholds Relative to the Severity Parameters of Items

All of the comparisons described to this point are based on thresholds for the probabilistic assignment of food security status that were specified to equate prevalence rates in the multiyear national sample to the corresponding rates based on discrete assignment of food security status. An advantage of probabilistic assignment is that thresholds could be specified to more precisely relate the ranges of severity to specific conditions as indicated by the cognitive content of items in the scale. Thresholds for discrete assignment are

Figure 5-8

Prevalence of food insecurity and very low food security based on discrete and probabilistic assignment by raw score on the 10-item adult scale, by race, Hispanic ethnicity, and income, average 1998-2006



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2006.

limited to integer values of raw score, and these may not always coincide precisely with conceptual specifications of measured conditions.

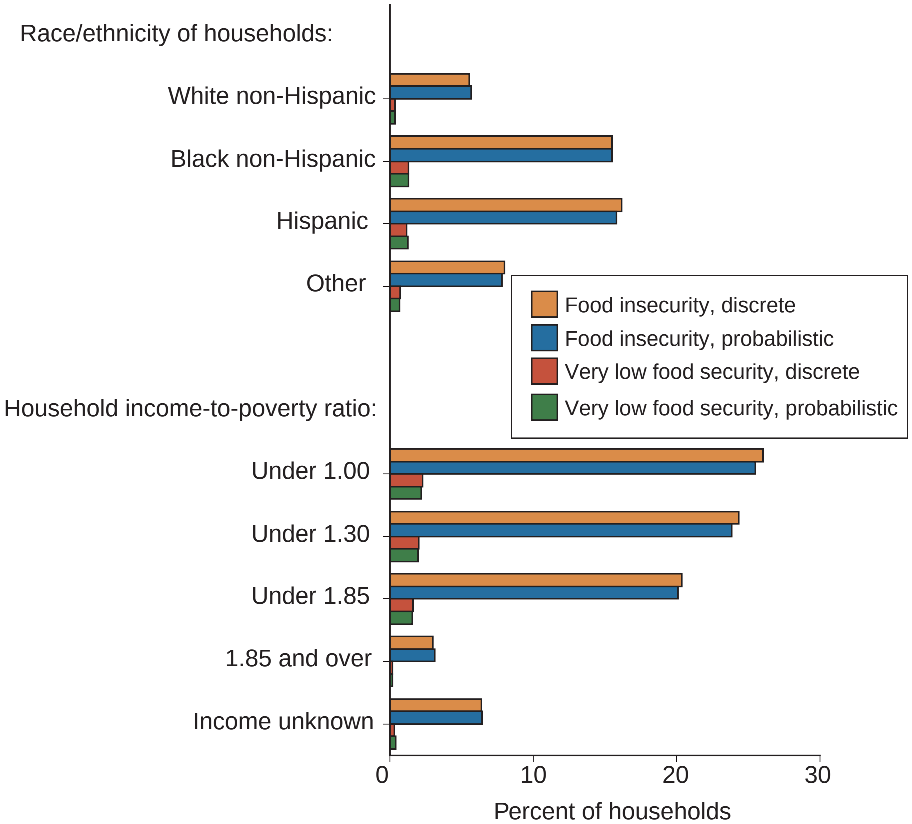
In this section, the prevalence rates that would be obtained by probabilistic assignment of food security status if the thresholds were set at the severity level of specific items in the respective scales are examined. For the household and adult scales, the selected items are those that were considered to be “threshold” indicators of food insecurity and very low food security (at that time described as “food insecurity with hunger”) by the team that originally developed the household food security scale (Hamilton et al., 1997a; 1997b). These were:

- For food insecurity, the household often or sometimes Could not afford to eat balanced meals.
- For very low food security, adults in the household cut the size of meals or skipped meals in 3 or more months because there was not enough money for food.

For the children’s food security scale, the selected items are those identified as threshold items in Nord and Bickel (2002). These were:

Figure 5-9

Prevalence of food insecurity and very low food security among children based on discrete and probabilistic assignment by raw score on the 8-item child scale, by race, Hispanic ethnicity, and income, average 1998-2006



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2006.

- For food insecurity among children (at that time described as "reduced quality and variety of children's diets"), the household often or sometimes could not afford to feed the children balanced meals.
- For very low food security among children (at that time described as "food insecurity with hunger among children"), at some time during the year the children were hungry but the household could not afford more food.

The current discrete threshold for food insecurity on the household scale may be slightly too low (not stringent enough) if the least severe condition that is considered to be appropriately considered food insecurity is the inability to afford balanced meals. The inconsistency is not very great, however. The estimated prevalence based on the current discrete method is 11.04 percent, compared with 10.71 percent based on a threshold set precisely at the severity level of the balanced meals item (table 5-2). The opposite is true of very low food security, and the difference is somewhat greater. The prevalence based on current discrete methods is 3.55 percent compared with 4.59 percent, based on a threshold set at the level of adults cutting the size of meals or skipping meals in 3 or more months.

The patterns for the adult scale are similar to those for the household scale.

Table 5-1

Standard errors of food security prevalence estimates based on discrete and probabilistic assignment of food security status, average 2004-06[1] and 2010[2]

Scale and estimate	Standard error based on discrete assignment of food security status	Standard error based on probabilistic assignment of food security status	Ratio of standard errors: probabilistic to discrete
	Percentage points		*Ratio*
Household scale			
Food insecurity, 2004-06 average	0.274	0.261	0.953
Food insecurity, 2010	.196	.181	.923
Very low food security, 2004-06 average	.161	.142	.884
Very low food security, 2010	.115	.099	.861
Adult scale			
Food insecurity among adults, 2004-06 average	.257	.242	.940
Food insecurity among adults, 2010	.183	.168	.918
Very low food security among adults, 2004-06 average	.159	.136	.857
Very low food security among adults, 2010	.118	.100	.847
Child scale			
Food insecurity among children, 2004-06 average	.333	.278	.836
Food insecurity among children, 2010	.280	.230	.827
Very low food security among children, 2004-06 average	.072	.049	.691
Very low food security among children, 2010	.089	.060	.674

[1]Standard errors for the 2004-06 estimates were calculated for each year separately using jackknife methods, then averaged across the three years.

[2]Standard errors for the 2010 estimates were calculated using balanced repeated replication methods based on replicate weights provided by the U.S. Census Bureau.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau Current Population Survey Food Security Supplements 2004-06 and 2010 for households with incomes less than 185 percent of the Federal poverty line.

Both of the current thresholds on the children's scale are slightly more stringent than thresholds at the severity level of the conceptually identified threshold items. The size of the difference in the case of food insecurity among children is not very great. About 7 percent more households with children would be identified as having food insecurity among children based on the item-related threshold (9.35 percent) than on the discrete threshold (8.76 percent). The difference in very low food security, however, may be considered substantial, as it amounts to about 23 percent (0.79 percent compared with 0.64 percent).

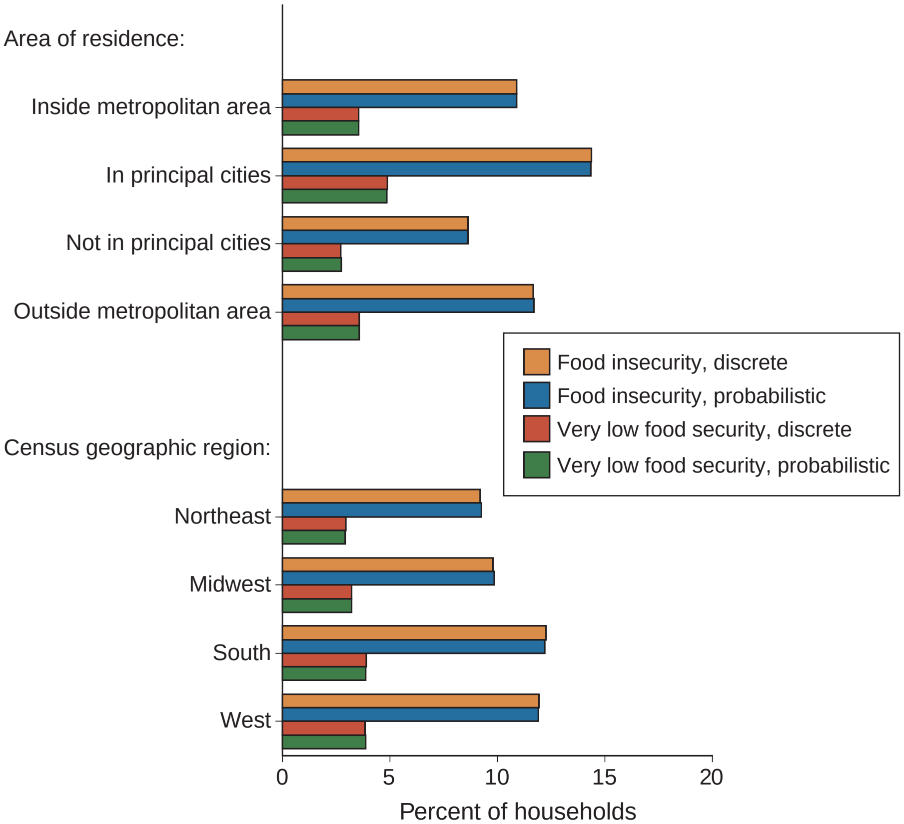
These comparisons do not imply that the current thresholds are wrong in some objective sense. An expert group might well determine that the appropriate threshold for a specific category of food insecurity is somewhere between the severity of two items, which could coincide precisely with a threshold based on a discrete raw score. The comparisons are intended to inform decisions on categorization procedures by characterizing the severity level of current thresholds in relation to the severity level of selected items.

Conclusions

With one exception, the discrete assignment of food security status has represented trends across time and comparisons across key population subgroups

Figure 5-10

Prevalence of food insecurity and very low food security based on discrete and probabilistic assignment by raw score on the 18-item household scale, by residence and Census region, average 1998-2006



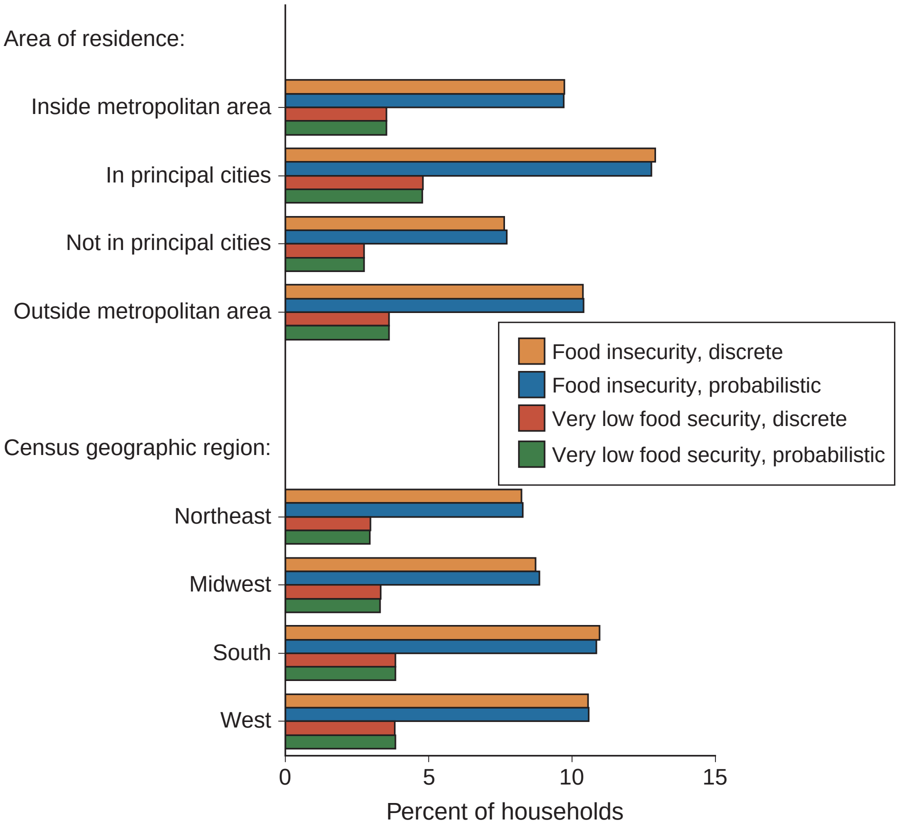
Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2006.

that are essentially undistorted compared with trends and comparisons based on probabilistic assignment of food security, which takes into account the measurement error inherent in the latent-trait measure. The exception is, however, an important one. The prevalence of food insecurity in households with children is overstated relative to households without children. Both methods show substantially higher prevalence of food insecurity in households with children, but the difference based on probabilistic assignment is about 65 percent while discrete assignment estimates the difference at about 100 percent (i.e., double).

Transparency and simplicity in high-visibility Federal Government measures of well-being are of great importance. Discrete assignment of food security status offers, in a sense, the best of both worlds. It allows the measure to draw on the scientific merits and statistical advantages of latent-trait measurement while supporting explanation of the measure to public and policy audiences based on raw score. The value of being able to say, “To be classified as food insecure, households must report at least these three conditions: ….” is enormous. USDA places considerable value on the public’s ability to replicate published prevalence estimates directly from the public-use data. The credibility of published estimates that results from this transparency is of great importance to USDA.

Figure 5-11

Prevalence of food insecurity and very low food security based on discrete and probabilistic assignment by raw score on the 10-item adult scale, by residence and Census region, average 1998-2006



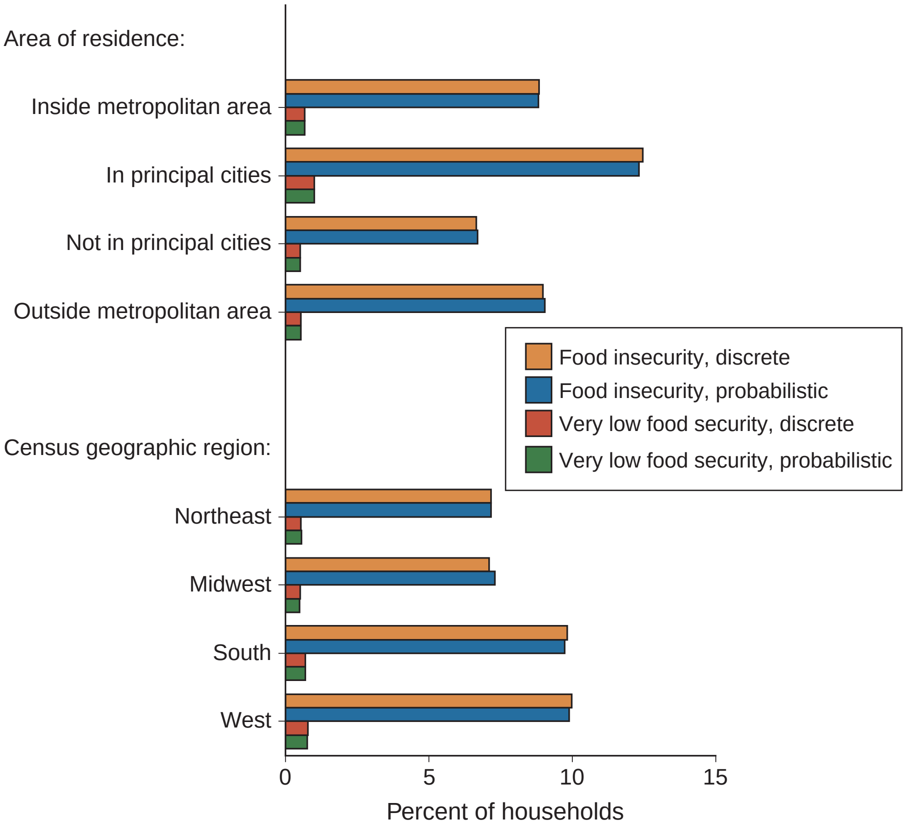
Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2006.

Considering the minimal distortion of all comparisons and trends except that between households with and without children, it is doubtful that the advantages of probabilistic assignment of food security status outweigh the advantages of discrete assignment.

The issue of comparisons between households with and without children will need to be addressed, but options other than probabilistic assignment of food security status may be preferable. One option is to continue to use discrete assignment based on the household scale and continue to describe the extent of the bias and the reason for it in a footnote. A problem is that the public and anti-hunger advocates tend to ignore the footnote and pick up only the “twice as high” comparison. A second option is to stop using the household food security scale as the primary monitoring method and, instead, use the adult and child scales (with food security status discretely assigned in each case) in combination. This is the approach Health Canada has adopted. Households are classified as food insecure if either adults or children or both are food insecure, and households are classified as having very low food security (described as “severe food insecurity” in the Health Canada system) if either adults or children or both are so classified. In the United States, too, the adult and child scales are already considered the primary measures for most research purposes because of the biases in the household scale associated with presence, absence, and age of children.

Figure 5-12

Prevalence of food insecurity and very low food security among children based on discrete and probabilistic assignment by raw score on the 8-item child scale, by residence and Census region, average 1998-2006



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2006.

Although these results provide little reason to use probabilistic assignment of food security status for calculating official prevalence statistics, they suggest that the methods could substantially improve estimates of sampling error. Application of probabilistic methods for this purpose could improve USDA's assessments of the statistical significance of differences in food security across subpopulations and over time. Similarly for some regression analyses, estimation errors may be reduced by modeling food insecurity as a probability instead of a discrete variable.

The probabilistic assignment methods have a number of other applications, and USDA has used these methods for several verification and assessment purposes in the past. For example, probabilistic methods were used in the 2000 CPS-FSS data to establish equivalent thresholds for the eight month-in-sample cases that had several nonstandard questions. The methods were used for a similar purpose in the 2007 CPS-FSS data, that is, to assess the extent of bias due to nonstandard versions of the "balanced meals" question that were tested, each in one-quarter of the sample. The methods have also been used to explore the extent of bias between households with and without children and to assess results from national surveys and research surveys with nonstandard food security question sets for comparability with published national statistics.

Table 5-2

National prevalence of household food insecurity and very low food security based on standard discrete assignment of food security status and on probabilistic assignment of food security status relative to thresholds set at the severity level of selected items

Scale Food security status	Prevalence based on discrete assignment of food security status[1]	Threshold item for probabilistic assignment of food security status	Threshold for probabilistic assignment of food security status[2]	Prevalence based on probabilistic assignment of food security status
	Percent		*Logistic units*	*Percent*
Household scale				
Food insecurity	11.04	Could not afford to eat balanced meals	4.343	10.71
Very low food security	3.55	Adult cut size of meals or skipped meals in 3 or more months	6.237	4.59
Adult scale				
Food insecurity among adults	9.84	Could not afford to eat balanced meals	4.343	10.67
Very low food security among adults	3.56	Adult cut size of meals or skipped meals in 3 or more months	6.237	4.60
Child scale				
Food insecurity among children	8.76	Could not afford to feed children balanced meals	5.386	9.35
Very low food security among children	.64	Children were hungry; could not afford more food	9.242	.79

[1]These are prevalence rates based on current standard methods. Each surveyed household is classified in a single food security status category based on raw score (and, for the household scale, presence or absence of children).

[2]Parameters for threshold items are from appendix table D-1.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2006 for households with incomes less than 185 percent of the Federal poverty line.

Chapter 6

Assessing Three Potential Enhancements in Combination

Abstract

This chapter assesses the combined effects of three potential technical improvements in USDA's food security measurement method. Earlier chapters have assessed each of these potential enhancements individually and two of them in combination. In this chapter, the performance of the most complex measurement model proposed—the 2-parameter logistic (2PL) model comprising polytomous items and using probabilistic assignment of food security status—is compared with that of the simplest model—the single-parameter logistic (1PL) model comprising dichotomous items and assigning food security status discretely. The findings provide no support for adoption of a polytomous 2PL model to measure food security. Prevalence rates across subpopulations of primary policy interest are distorted minimally, if at all, by the current simpler measure. Gains in reliability of measured severity at the household level would be small or zero. The more complex model might improve the precision of prevalence estimates of very low food security modestly (results differ between the two time periods), but at the cost of lower precision for estimates of food insecurity.

Background

The Committee on National Statistics (CNSTAT) panel that reviewed USDA's food security measurement methods suggested several ways in which the methodology might be refined, contingent on confirmatory research (NRC, 2006). Earlier chapters have assessed five of those potential methodological enhancements. In this chapter, three of the potential enhancements are assessed in combination. The performance of the simplest scale—the 1PL dichotomous model using discrete assignment of food security status—is compared with that of the most complex model—the 2PL polytomous scale using probabilistic assignment of food security status. The latter model relaxes the constraint that all items have equal discrimination, includes responses to followup questions on how frequently food-insecure conditions occurred, and takes measurement error into account by assigning food security status as a probability of food insecurity. The performance of these models is also compared with the performance of several intermediate models to clarify which of the differences in modeling account for any observed difference in performance.

The three potential enhancements examined in combination in this study, and key findings from the previous assessments of each of them individually are as follows:

- (Chapter 1)—Incorporating into the measure all of the frequency-of-occurrence information routinely collected in the CPS-FSS: Analysis suggested that this may not be appropriate because frequent/chronic severity appears to represent a somewhat distinct dimension from ever-during-the-year severity. Furthermore, the 1PL polytomous scale using discrete assignment of food security status would provide small to minimal gains in measurement precision.

- (Chapter 3)—Using a 2PL measure, which would allow item discrimination to differ between items: Analysis suggested that although 2PL analysis may contribute to item and scale assessment, a measure based on the 2PL model would provide negligible improvement in the applications of most importance and visibility in USDA's food security monitoring system.
- (Chapter 5)—Taking measurement error into account by basing prevalence rates on the probability of each surveyed household's being food insecure: Analysis suggested that, with the important exception of the difference between households with and without children, this method would have minimal effect on comparisons of prevalence rates of food insecurity over time or across subpopulations. However, prevalence rates based on probabilistic assignment would probably have smaller errors due to measurement error at the household level than those based on the current method of discrete assignment of food security status.

In Chapter 3, the latter two methods were also assessed in combination, with results similar to those described for each method assessed individually. The "full-information" polytomous model examined in Chapter 1 was not assessed in combination with either of the other two methods.

A 2PL model with probabilistic assignment of food security status is, in principle, a more precise and reliable representation of the underlying latent trait of food insecurity than a 1PL model based on the same items. It can be assumed, therefore, that the simpler 1PL dichotomous model misrepresents the underlying latent trait to some extent. However, the 1PL model with discrete assignment of food security status has considerable advantages in high-visibility public statistics because of its transparency, ease of communication, and ease of implementation in surveys across the national food security monitoring system and in State and local monitoring and academic research applications. USDA prefers, therefore, to base its published food security statistics on the simpler model unless doing so distorts conditions or trends in food security to a substantively important extent.

It is less clear whether the 2PL polytomous model, which incorporates additional information of frequency of occurrence, is an improvement over the 2PL dichotomous model. The assessment of the 1PL polytomous model in Chapter 1 raised questions about the extent to which severity and frequency of food insecurity are distinct dimensions. Whether for that reason or because respondents report frequency of occurrence somewhat inconsistently, the 1PL polytomous model provided little or no improvement in measurement error compared with the current standard dichotomous scale.

The issues addressed in this chapter, then, are: (1) the extent to which the current, simple model distorts USDA's representation of food insecurity across key subpopulations, and (2) the extent to which a more technically correct model would improve the reliability of the measure (i.e., reduce measurement error) compared with the simple model currently in use. Prevalence rates of food insecurity and very low food security across subpopulations for which prevalence rates are routinely reported by USDA are compared based on the simplest and most complex measures. Finally, the reliability of those measures are compared, and the sizes of estimation errors on national prevalence rates of food insecurity and very low food security are compared across measures.

Data and Methods

Data

The main analyses used low-income subsamples of the nationally representative CPS-FSS data for 2003, 2004, and 2005. The large size of the 3-year sample reduces uncertainties due to sampling. Data from the December 2010 CPS-FSS were used in a parallel analysis to take advantage of the improved capability these data offer to assess measurement error using balanced repeated replication (BRR) methods, The 2010 CPS-FSS was the first Food Security Supplement for which the U.S. Census Bureau calculated replicate weights to support estimation of sampling errors using BRR methods.

The samples were restricted to households with incomes less than 185 percent of the poverty line to avoid distortions that could arise from additional screening procedures that are applied to higher income households. Higher income households are screened out of the food security questions entirely unless they give an indication of some level of food-access difficulties on either of two preliminary screeners. Households with missing responses to one or more of the household or adult food security questions or the frequency-of-occurrence followup questions (a small proportion of households) were omitted. The resulting samples consisted of 34,911 households in 2003-05 and 11,323 households in 2010. Samples for estimating measurement model parameters omitted households that denied all 8 items, leaving 15,178 households in 2003-05 and 5,565 households in 2010. Samples for estimating the CML models also omitted the small proportions of households that affirmed all 8 items, leaving 14,607 households in 2003-05 and 5,336 households in 2010.

Measures of food insecurity

Measures of adult food security based on five models were analyzed for each time period.

1. 2PL polytomous scale. The first three items—*Worried food would run out, Food bought did not last,* and *Could not afford to eat balanced meals*—were coded as trichotomies: never/sometimes/often. The next three items—*Adult cut size of meals or skipped meals, Respondent ate less than should*, and *Respondent hungry*—were coded as four-category polytomies: never/in-only-1-or-2-months/some-months-but-not-every-month/almost-every-month. The item—*Lost weight*—was coded as a dichotomy: no/yes. The final item—*Adult did not eat whole day*—was coded as a trichotomy: never/in-only-1-or-2-months/some-months-but-not-every-month-or-almost-every-month. (The two categories indicating highest frequency of occurrence in this item were combined to avoid nonconvergence problems in the estimation. The almost-every-month category for this item was reported by only a very few households, almost

all with extreme responses.) The model was estimated using marginal maximum likelihood (MML) methods implemented by Parscale 4 software (Scientific Software International, 2003). The intercategory structure of each item was estimated separately, not constrained to be equal across items. Respondent measures and errors were the expected a posteriori (EAP) estimates.

2. 1PL polytomous scale. This was identical to model 1 except that the discrimination of all items was constrained to be equal. Intercategory distances were not constrained to be equal across items. The model was estimated using MML methods implemented by Parscale 4, with respondent measures and errors based on EAP estimates.

3. 2PL dichotomous scale. This scale included the same eight items as models 1 and 2, but each was coded as a dichotomy: never/ever. The model was estimated using MML methods implemented by Parscale 4 with respondent measures and errors based on EAP estimates.

4. 1PL dichotomous scale, MML with EAP respondent parameters. This was the same as model 3 except that the discrimination of all items was constrained to be equal. The model was estimated using MML methods implemented by Parscale 4 with respondent measures and errors based on EAP estimates.

5. 1PL dichotomous scale, CML with maximum likelihood (ML) respondent parameters based on CML item probabilities in each raw score. This was identical to model 4 except that respondent severity parameters were ML estimates calculated from CML-estimated item-response probabilities, and respondent errors were estimated as the inverse square root of the Fisher information function. These are the primary methods USDA uses currently to assess food security item and scale performance. This model was estimated following procedures described by Fischer and Molenaar (1995) using programs developed by the author and implemented in SAS data steps.[1]

The performance of each scale was assessed with probabilistic assignment of food security status, and 1PL dichotomous scale (measure 4 or 5) was assessed with discrete assignment of food security as a baseline for comparison.

Discrete assignments of food security status based on models 4 and 5 are identical. Both are consistent with current USDA methods, except that two items in the current USDA scale, *adult cut size of meal or skipped meal* and *adult did not eat whole day*, are trichotomies to include frequency-of-occurrence information coded as described for the latter item in model 1. In the analyses for this study, those items were coded as dichotomies to more clearly contrast the polytomous and dichotomous scales. Models 4 and 5, even though identical in terms of discrete assignment and essentially identical in terms of respondent severity parameters, differ in the size of their error estimates, and so differ in the probabilities of food security they assign to households in each raw score.

Models 2 and 3 represent alternative intermediate “steps” between the simplest models (models 4 and 5) and the most complex model (model 1).

[1]The CML estimation programs were developed by the author and have been found to give identical results when tested against the commercially available Winmira software and against SAS Proc Logistic with the “Strata” option.

Comparisons of these intermediate models with models 1 and 4 (or 5) will shed light on whether any differences between the simplest and most complex models are due to relaxing the equal discrimination constraint or to the additional information in the frequency-of-occurrence followup responses,

Discrete and Probabilistic Assignment of Food Insecurity and Very Low Food Security

Discrete assignment for the 1PL dichotomous models was as follows: Households with raw scores of 3 or higher were classified as food insecure (including low and very low food security). Households with raw scores 6 or higher were classified as having very low food security. These replicate the USDA official classifications as nearly as possible without including the frequency of cutting or skipping meals.

Each household was also assigned a probability of food insecurity and very low food security on each scale, based on the household's severity parameter and estimated measurement error on that scale.[2] Probabilities of food insecurity on each measure were calculated as follows:

- An initial threshold value on the latent trait was selected so as to approximately equate the mean probability of food insecurity across all households to the proportion classified as food insecure by discrete assignment by raw score on the 1PL dichotomous scale.
- The probability of food insecurity for each household was calculated as the proportion of a normal distribution beyond the threshold, with the mean of the distribution equal to the household's severity-parameter estimate and standard deviation equal to the measurement error.
- The population estimate was calculated as the weighted mean of probabilities across households.
- The initial threshold was then adjusted iteratively until the population estimate matched the target—the proportion classified as food insecure based on discrete assignment by raw score on the 1PL dichotomous scale.

This calculation was repeated for food insecurity and very low food security for each measure.

In the calculations of equivalent thresholds, households with zero raw score were included and considered to be fully food secure with no measurement error, and households with maximum raw score were included and considered to have very low food security with no measurement error. Measurement error cannot be calculated for extreme scores in the CML model and cannot be calculated in MML models without making assumptions about the distribution of the latent trait. Assumptions about that distribution for households with raw score zero would be of dubious value and difficult to justify because such households comprise 40 to 60 percent of all households in these samples.

Comparing Prevalence Rates Across Subpopulations

Prevalence rates of food insecurity and very low food security based on the simplest and most complex models were compared across subpopulations by

[2]Alternatively, the probability of food insecurity for the interviewed household may be viewed as the estimated proportion of households in the population (represented by that interviewed household) that are food insecure.

household composition, race and Hispanic ethnicity, income relative to the poverty line, residence (relative to metropolitan statistical areas), and Census geographic region. Classification of households by these characteristics was based on data in the Current Population Survey core following the methods used in USDA's annual food security reports. These comparisons used the 2003-05 data.

Assessing Measurement Reliability Across Models

Comparisons of measurement error between models were not straightforward because it was not clear how to adjust the measures to equivalent metrics. Comparing Rasch reliability statistics avoided that problem. The Rasch reliability statistic is calculated by decomposing the total variance—the sum of squared differences of true household values on the latent trait from the grand mean—into the variance accounted for by the model and the error variance. Total variance for each model was calculated as sum of two components: (1) the weighted variance across unique response patterns (across raw scores for the 1PL models), and (2) the mean of the squared measurement error across households. Rasch reliability is the variance across unique response patterns (across raw scores for the 1PL models) as a proportion of total variance. Both variance components were calculated based on survey weights and included only cases nonextreme on the dichotomous models in order to be comparable.[3]

Higher Rasch reliability does not necessarily result in similarly higher classification reliability, however. Rasch reliability is calculated across the entire measured range, but higher reliability in a range of the latent trait far from a threshold may have little impact on classification vis-à-vis that threshold. Thus, although Rasch reliability provides information on potential improvement that may result from a more complex measure, an assessment of classification measurement error provides information on realization of that potential.

Comparing Measurement Error on Food Security Prevalence Estimates Across Models

The primary application of the food security scale in the realm of public statistics is to classify households with regard to their food security status. The food security status of the population and of selected subpopulations is then summarized by the prevalence of food insecurity at various levels of severity as represented by those classifications (see, for example, Coleman-Jensen et al., 2011).

Thus, the “bottom line” assessment of a more complex model is the extent to which it would improve the accuracy of USDA's annual prevalence estimates of food insecurity and very low food security in U.S. households compared with a simpler model. Replication-based estimates of errors in prevalence statistics directly assess this most important practical effect of differences in measurement error between measures.

Errors of estimated prevalence rates of food insecurity and very low food security based on discrete assignment of food security status by the 1PL discrete model (measure 4 or 5) and on probabilistic assignment by each of

[3]Rasch reliability can be considered to compare measurement precision after equating the population standard deviation of the latent trait as estimated by the respective measure.

the five measures were calculated using two replication methods. Jackknife replication methods were used for the 2003-05 data and BRR methods for the 2010 data. Both of these methods are designed for the primary purpose of estimating sampling error, but in fact, both methods estimate the combined sampling error and measurement error. Indeed, the two sources of error cannot be differentiated by these methods. For each data set used in this study, sampling error was the same for all measurement methods, since they used the same data, so differences across methods in replication-estimated errors can be ascribed to differences in measurement error.

The jackknife estimates used the eight month-in-sample "rotation" groups in each year, since these are, effectively, independent samples of U.S. households. Standard jackknife replication methods were applied to these subsamples to calculate error estimates for prevalence rates of food insecurity and very low food security for each of the measures for the 3-year period 2003-05.[4]

The BRR methods using the December 2010 CPS-FSS data were implemented in SAS Proc SurveyMeans using replicate weights provided by the Census Bureau.[5]

Findings

Prevalence rates for subpopulations estimated by the most complex model differed little from those estimated by the simplest model (figs. 6-1, 6-2, and 6-3). Differences in prevalence rates between the two methods were small relative to the differences across the subpopulations. The largest between-method differences, proportionally, were higher rates of very low food security among elderly, based on the more complex model compared with the simplest model. Findings in Chapter 1 suggest that this results from modeling the items as polytomies reflecting the frequency or persistence of food insecurity. Because economic stability is greater among elderly, those with very low food security are more likely to remain in that condition persistently.

Reliability of the most complex model (polytomous 2PL) was higher than that of the simplest model (dichotomous 1PL) when the latter was also estimated in MML, but not when it was estimated in CML (table 6-1). Since the household severity parameters of the MML and CML estimates of the dichotomous 1PL model are almost perfectly collinear (Pearson correlation > .999, analysis not shown), the difference in reliability between them reflects a difference in their error estimates. While it is not certain which method estimates the errors more accurately, the MML estimates were based on the same methodology as the estimates for the polytomous 2PL model. So it appears that the more complex model provides more reliable estimates. The reliability of the polytomous 1PL model was at least as good as that of the polytomous 2PL model in both years, so the gain in reliability appears to have been primarily due to including the frequency-of-occurrence information, rather than to relaxing the constraint that all items have equal discrimination.

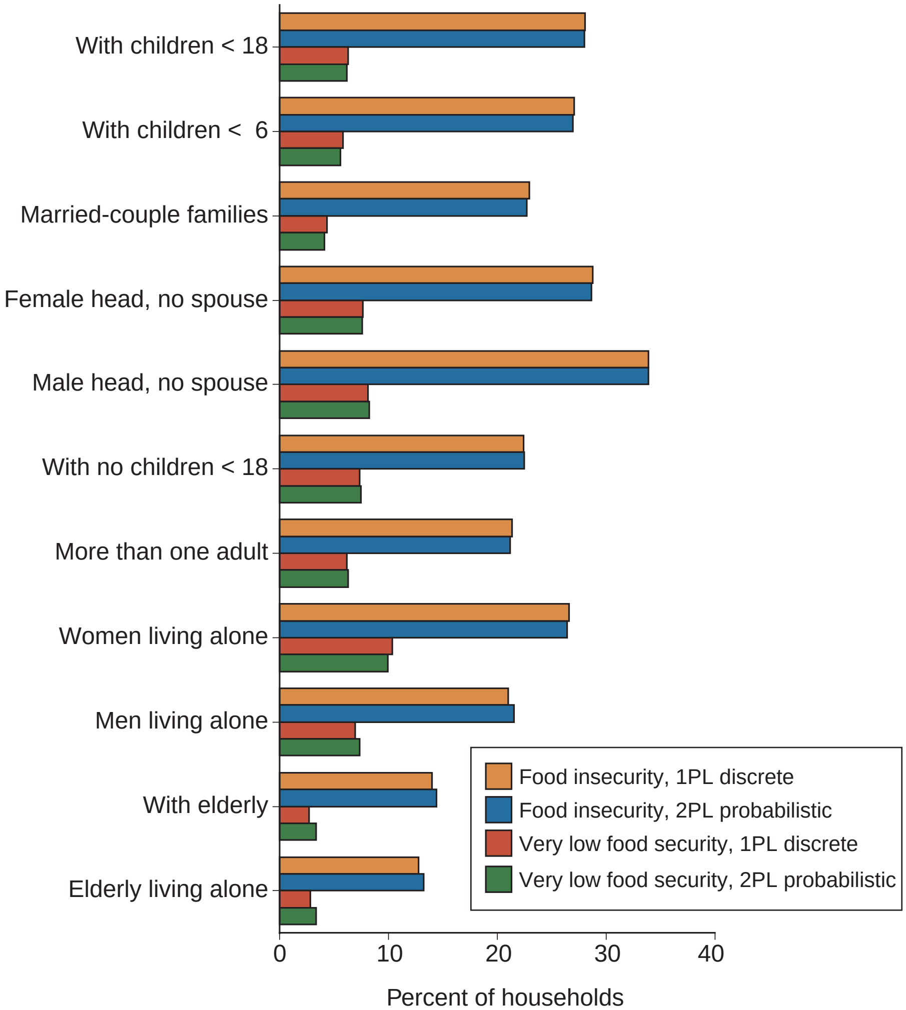
However, the higher reliability does not consistently result in reduced estimation errors for prevalence rates of food insecurity and results in, at most, small reductions in estimation errors for prevalence rates of very low food security. Estimation error (i.e., combined sampling and measurement error)

[4]To avoid treating repeated interviews of households in 2 subsequent years as independent, the month-in-sample groups with potential re-interviews were grouped in the same jackknife subsample.

[5]Replicate weights were first provided for the CPS-FSS in 2010. The weights are based on the Fay method and, following specifications provided by the Bureau, the option "Fay=.5" was specified in the SAS SurveyMeans procedure.

Figure 6-1

Prevalence of food insecurity and very low food security based on 1-parameter logistic (1PL) scale of dichotomous items using discrete assignment of food security status and 2-parameter logistic (2PL) scale of polytomous items using probabilistic assignment of food security status, by household composition, average 2003-05



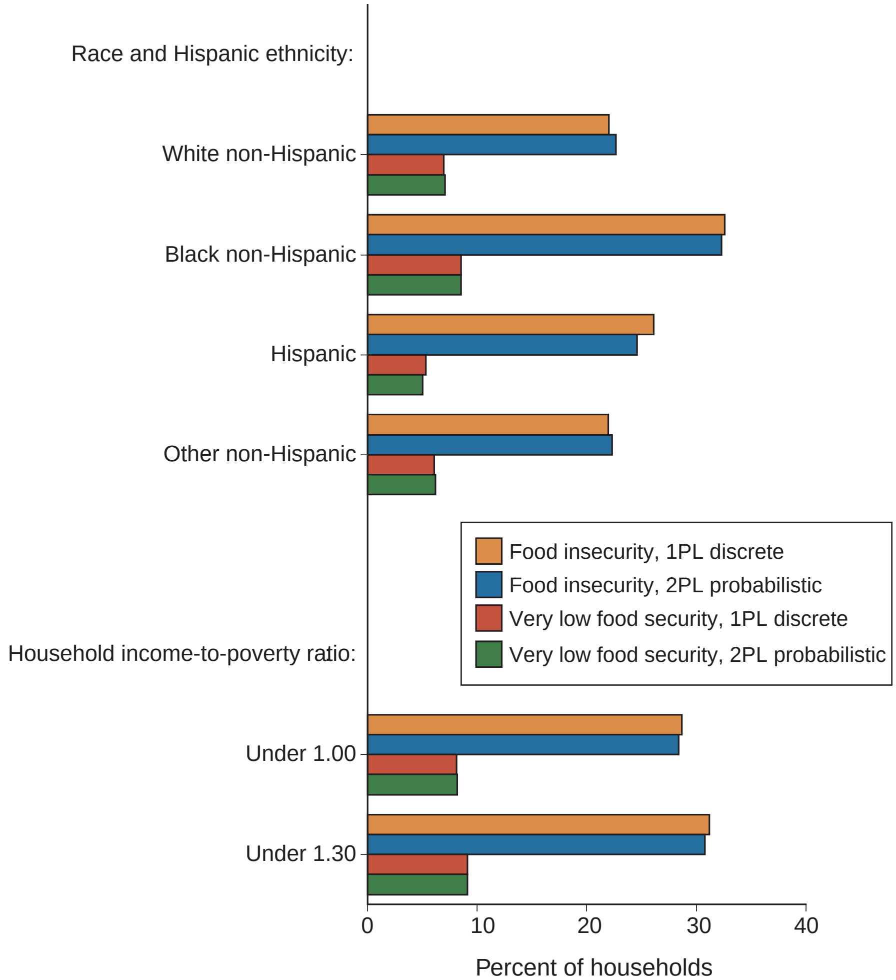
Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2003, 2004, and 2005 for households with incomes less than 185 percent of the Federal poverty line.

for the prevalence of food insecurity was larger for any of the more complex models than for the dichotomous 1PL model (table 6-2). For example, in 2003-05, estimation error on the prevalence rate of food insecurity was 0.398 percentage points for the polytomous 2PL model with probabilistic assignment, compared with 0.353 percentage points for the dichotomous 1PL model, or 10.8 percent larger (ratio 1.108) for the more complex model. In the 2010 data, using the methodologically stronger BRR variance estimation, errors differed little across the measures.

For very low food security, the picture was more complex. In the 2003-05 data, estimation errors were larger or essentially the same for the more complex models as for the dichotomous 1PL model. However, the opposite was true in the 2010 data. Errors for the polytomous models were 4 to 6

Figure 6-2

Prevalence of food insecurity and very low food security based on 1-parameter logistic (1PL) scale of dichotomous items using discrete assignment of food security status and 2-parameter logistic (2PL) scale of polytomous items using probabilistic assignment of food security status, by race, Hispanic ethnicity, and by income, average 2003-05



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2003, 2004, and 2005 for households with incomes less than 185 percent of the Federal poverty line.

percent smaller (ratios .958 and .942) than for the dichotomous 1PL model with errors estimated by CML.

If a more complex model were to be considered, it is worth noting that there is little or no reason to prefer a 2PL model over a 1PL model. In all eight comparisons in table 6-2, the 1PL model produced smaller measurement errors than the otherwise similar 2PL model.

The comparison of estimation errors between the CML and MML dichotomous 1PL models shed some light on the issue of the difference in their calculated reliabilities described earlier. With probabilistic assignment of food security status, prevalence estimation errors for very low food security were slightly smaller for the MML model than for the CML model. For food insecurity, estimation errors were essentially the same for the two models.

Table 6-1
Reliability of food security measures

Model	Reliability[1]	
	2003-05 data	2010 data
Polytomous 2-parameter (2PL)	.734	.766
Polytomous 1-parameter (1PL)	.741	.768
Dichotomous 2PL	.725	.722
Dichotomous 1PL (MML)	.679	.729
Dichotomous 1PL (CML)	.764	.773
Number of households	14,607	5,336

CML = Conditional maximum likelihood.
MML = Marginal maximum likelihood.

Note: The polytomous models differentiated affirmative responses into two or three categories depending on how often the condition or behavior occurred. The dichotomous models collapsed all affirmative responses into a single category.

[1]Reliability was calculated by decomposing the total variance—the sum of squared differences of true household values on the latent trait from the grand mean—into two components: (1) the variance accounted for by the model, calculated as the weighted variance across unique response patterns (or across raw scores for the 1PL models), and (2) the mean of the squared measurement error across households. Reliability is the variance across unique response patterns (or raw scores) as a proportion of total variance. Both variance components were calculated based on survey weights, and included only cases nonextreme on the dichotomous models in order to be comparable.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau Current Population Survey Food Security Supplements for 2003, 2004, 2005, 2010 for households with incomes less than 185 percent of the Federal poverty line.

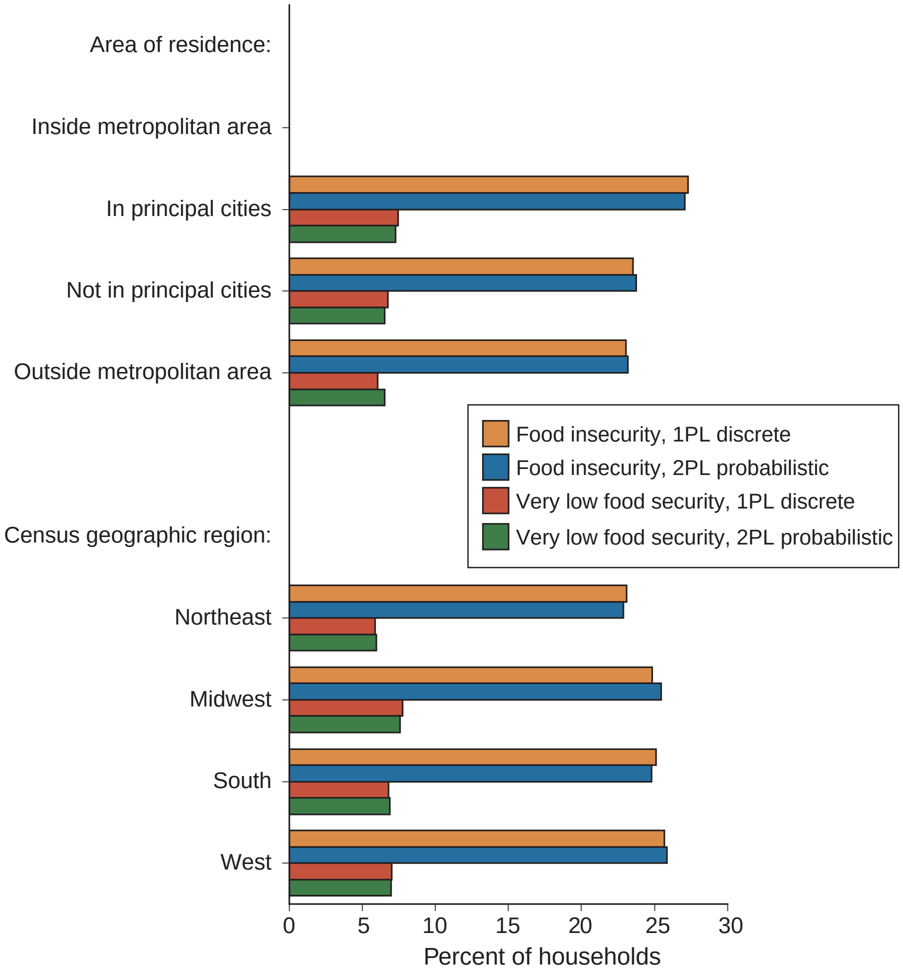
This suggests that the measurement error of severity *at the household level* is more accurately estimated in the MML model than in the CML model.

Conclusions

The findings provide little or no support for adoption of a polytomous 2PL model to measure food security. Prevalence rates across subpopulations of primary policy interest are distorted minimally, if at all, by the current simpler measure. Gains in reliability of measured severity at the household level would be small or zero. The polytomous 2PL model might improve the precision of prevalence estimates of very low food security modestly (results differ between the two time periods), but at the cost of lower precision for estimates of food insecurity. The extent of improvement in precision that might be realized is not likely to justify the loss of transparency, simplicity, and communicability that would result from use of the more complex model.

Figure 6-3

Prevalence of food insecurity and very low food security based on 1PL scale of dichotomous items using discrete assignment of food security status and 2PL scale of polytomous items using probabilistic assignment of food security status, by residence and Census region, average 2003-05



1PL = 1-parameter logistic.
2PL = 2-parameter logistic.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 2003, 2004, and 2005 for households with incomes less than 185 percent of the Federal poverty line.

Table 6-2

Combined sampling and measurement error on national prevalence rates of food insecurity and very low food security based on five measurement models

	Error (percentage points)[2]		Ratio to dichotomous 1PL estimated in CML	
	2003-05	2010	2003-05	2010
Food insecurity – discrete assignment (prevalence: 24.89 percent in 2003-05 and 30.61 percent in 2010)[1]				
Dichotomous 1PL (CML or MML)	.353	.496		
Food insecurity – probabilistic assignment				
Polytomous 2PL	.398	.453	1.108	1.027
Polytomous 1PL	.381	.442	1.060	1.001
Dichotomous 2PL	.386	.461	1.075	1.044
Dichotomous 1PL (MML)	.358	.438	.995	.992
Dichotomous 1PL (CML)	.360	.441	Ref.	Ref.
Very low food security – discrete assignment (prevalence: 24.89 percent in 2003-05 and 30.61 percent in 2010)[1]				
Dichotomous 1PL (CML or MML)	.233	.321	Ref.	Ref.
Very low food security – probabilistic assignment				
Polytomous 2PL	.245	.255	1.062	.958
Polytomous 1PL	.242	.250	1.047	.942
Dichotomous 2PL	.226	.263	.981	.990
Dichotomous 1PL (MML)	.225	.256	.976	.964
Dichotomous 1PL (CML)	.231	.266	Ref.	Ref.

[1]Errors for prevalence estimates were calculated by Jackknife replication methods based on month-in-sample in the 2003-05 data and by balanced repeated replication (BRR) methods using replicate weights provided by the Census Bureau in the 2010 data. Both methods inherently estimate combined sampling and measurement error.

[2]Prevalence rates are based on the dichotomous 1PL model with discrete assignment of food security status, the simplest model and most similar to USDA's current standard measure. Thresholds for probabilistic assignment exactly replicated the target prevalence rates.

1PL = 1-parameter logistic; 2PL = 2-parameter logistic.
CML = Conditional maximum likelihood.
MML = Marginal maximum likelihood.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau Current Population Survey Food Security Supplements for 2003, 2004, 2005 for households with incomes less than 185 percent of the Federal poverty line.

References

Bickel, Gary, Mark Nord, Cristofer Price, William Hamilton, and John Cook. 2000. *Guide to Measuring Household Food Security, Revised 2000*. USDA, Food and Nutrition Service. Available at: http://www.fns.usda.gov/fsec/files/fsguide.pdf

Bond, Trevor G., and Christine M. Fox. 2001. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahway, NJ: Lawrence Erlbaum Associates, Publishers.

Coleman-Jensen, Alisha, Mark Nord, Margaret Andrews, and Steven Carlson. 2011. *Household Food Security in the United States in 2010*. ERR-125, USDA, Economic Research Service. Available at: http://www.ers.usda.gov/publications/err-economic-research-report/err125.aspx

Fischer, Gerhard H., and Ivo W. Molenaar, eds. 1995. *Rasch Models: Foundations, Recent Developments, and Applications*. New York: Springer-Verlag.

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*. Prepared for USDA, Food and Consumer Service. Available at: http://www.fns.usda.gov/oane/menu/published/foodsecurity/sumrpt.pdf

Hamilton, W.L., J.T. Cook, W.W. Thompson, L.F. Buron, E.A. Frongillo, Jr., C.M. Olson, and C.A. Wehler. 1997b. *Household Food Security in the United States in 1995: Technical Report*. Prepared for USDA, Food and Consumer Service. Available at: http://www.fns.usda.gov/oane/menu/published/foodsecurity/tech_rpt.pdf

Masters, G.N. 1982. “A Rasch model for partial credit scoring,” *Psychometrika* 47: 149-174.

Mislevy, Robert J. and R. Darrell Bock. 1990. *Bilog 3: Item Analysis and Test Scoring with Binary Logistic Models*. Mooresville, IN: Scientific Software, Inc.

National Research Council. 2006. *Food Insecurity and Hunger in the United States: An Assessment of the Measure*. Committee on National Statistics, Panel to Review the U.S. Department of Agriculture’s Measurement of Food Insecurity and Hunger, Gooloo S. Wunderlich and Janet L. Norwood (eds.). Washington, DC: The National Academies Press.

Nord, Mark. 2006. *Exploring Item Fit Statistics and Measurement Reliability of Food Security Scales Using Simulated Data*, paper presented at the annual meeting of the American Agricultural Economics Association, Long Beach, CA, July 23-26.

Nord, Mark, and Gary Bickel. 2002. *Measuring Children’s Food Security in U.S. Households, 1995-99*. FANRR- 25, USDA, Economic Research Service. Available at: http://webarchives.cdlib.org/sw15d8pg7m/http://ers.usda.gov/publications/fanrr25

Nord, Mark and C. Philip Brent. 2002. *Food Insecurity in Higher Income Households*. E-FAN 02016, USDA, Economic Research Service. Available at http://webarchives.cdlib.org/sw1q52gj15/http://www.ers.usda.gov/Publications/EFAN02016

Nord, Mark and Debra A. Fogarty. 2000. *Comparing Prevalence Rates of Food Insecurity and Hunger Based on Non-Standard Question Sets to the National Benchmark*, presentation at the Fourth International Conference on Dietary Assessment Methods, Tucson, AZ, September 17-20.

Nord, Mark and John M. Linacre. 2006b. *Improving Factor Analytic Assessment of Residuals in Rasch-Model-Based Scales*, paper presented at the annual meeting of the American Agricultural Economics Association, Long Beach, CA, July 23-26.

Nord, Mark, and Kathy Radimer. 2005. *Measuring Frequent or Chronic Food Insecurity in Household Surveys,* paper presented at the annual research conference of the Federation of American Societies for Experimental Biology, San Diego, CA, April 2-6.

Ohls, James, Larry Radbill, and Allen Schirm. 2001. *Household Food Security in the United States, 1995-1997: Technical Issues and Statistical Report*. Prepared by Mathematic Policy Research, Inc., for USDA, Food and Nutrition Service. Available at: http://www.fns.usda.gov/oane/MENU/Published/FoodSecurity/FoodSecurityTech.pdf

Opsomer, Jean D., Helen H. Jensen, and Suwen Pan. 2003. “An Evaluation of the U.S. Department of Agriculture Food Security Measure with Generalized Linear Mixed Models,” *Journal of Nutrition* 133: 421-427.

Raju, Nambury S. 1988. “The Area between Two Item Characteristic Curves,” *Psychometrika* 53(4): 495-502.

Ryu, Jeong-Hee, and Judi Bartfeld. 2011. *Long-Term Food Security Patterns and Children’s Cognitive and Health Outcomes*, final report on cooperative research project, “Dynamics of food insecurity using the Early Childhood Longitudinal Survey-Kindergarten Cohort (ECLS-K),” USDA, Economic Research Service.

Scientific Software International, Inc. 2003. Parscale for Windows, Version 4.1.2328.4. Lincolnwood, IL: Scientific Software, Inc.

Wilde, Parke E. 2004. “Differential Response Patterns Affect Food-Security Prevalence Estimates for Households with and without Children,” *Journal of Nutrition* 134: 1920-1915.

Wilde, Parke E., Mark Nord, and Robert E. Zager. 2010. “In Longitudinal Data from the Survey of Program Dynamics, 16.9% of the U.S. Population Was Exposed to Household Food Insecurity in a Five-Year Period,” *Journal of Hunger and Environmental Nutrition* 5(3): 380-398.

Wright, B.C. and L.M. Linacre. 1994. “Reasonable Mean-Square Fit Values,” *Rasch Measurement Transactions* 8(3):370. Available: http://www.rasch.org/rmt/rmt83.htm

Appendix A—Questions About Household- and Adult-Level Food Security in the Current Population Survey Food Security Supplement

1. "We worried whether our food would run out before we got money to buy more." Was that often, sometimes, or never true for you in the last 12 months?[1]

2. "The food that we bought just didn't last and we didn't have money to get more." Was that often, sometimes, or never true for you in the last 12 months?

3. "We couldn't afford to eat balanced meals." Was that often, sometimes, or never true for you in the last 12 months?

4. In the last 12 months, did you or other adults in the household ever cut the size of your meals or skip meals because there wasn't enough money for food? (Yes/No)

    4a. (If yes to Question 4) How often did this happen—almost every month, some months but not every month, or in only 1 or 2 months?

5. In the last 12 months, did you ever eat less than you felt you should because there wasn't enough money for food? (Yes/No)

    5a. (If yes to Question 5) How often did this happen—almost every month, some months but not every month, or in only 1 or 2 months?

6. In the last 12 months, were you ever hungry, but didn't eat, because you couldn't afford enough food? (Yes/No)

    6a. (If yes to Question 6) How often did this happen—almost every month, some months but not every month, or in only 1 or 2 months?

7. In the last 12 months, did you lose weight because you didn't have enough money for food? (Yes/No)

8. In the last 12 months did you or other adults in your household ever not eat for a whole day because there wasn't enough money for food? (Yes/No)

    8a. (If yes to Question 8) How often did this happen—almost every month, some months but not every month, or in only 1 or 2 months?

[1]In the interview, alternative fills such as (I/we), (you/your household), (you/you or other adults in your household) are selected by the computer-assisted interviewing software to be consistent with household composition.

Appendix B–Item Parameters With Dependent Item-Pairs Modeled as Trichotomies, and Food Security Status Classifications Based on Raw Scores

Item parameters for the three scales examined in Chapter 2, estimated from the 2003-05 data, are shown in table B-1. The models for the child and adult scales were constrained to have the same mean and standard deviation of item parameters as those of the same items in the household scale in order to facilitate comparison. The discrimination parameters required to equate the standard deviations are also tabled.

Food security status classifications on the three scales are as follows:

Household Scale (for households with children)

Raw score 0 – High food security

Raw score 1-2 – Marginal food security

(The combined categories of high and marginal food security are referred to jointly as food security)

Raw score 3-7 – Low food security

Raw score 8-18 – Very low food security

(The combined categories of low and very low food security are referred to jointly as food insecurity.)

Adult Scale (and household scale for households with no children)

Raw score 0 – High food security among adults

Raw score 1-2 – Marginal food security among adults

(The combined categories of high and marginal adult food security are referred to jointly as adult food security.)

Raw score 3-7 – Low food security among adults

Raw score 8-18 – Very low food security among adults

(The combined categories of low and very low adult food security are referred to jointly as adult food insecurity.)

Child Scale

Raw score 0-1 – Food security among children

(High and marginal food security cannot be fully differentiated on the child scale.)

Raw score 2-4 – Low food security among children

Raw score 8-18 – Very low food security among children

(The combined categories of low and very low food security among children are referred to jointly as food insecurity among children.)

Table B-1

Item-severity parameter estimates, 15-item household scale and 7-item child scale

	Household scale		Adult scale[1]		Child scale[1]	
	Rasch parameter	Dichotomous parameter.[2]	Rasch parameter	Dichotomous parameter.[2]	Rasch parameter	Dichotomous parameter.[2]
Worried food would run out	2.41		2.62			
Food bought did not last	3.63		3.70			
Could not afford to eat balanced meals	4.45		4.15			
Few kinds of low-cost food for children	3.87				3.49	
Could not feed children balanced meals	5.49				5.31	
Children not eating enough	6.99				7.08	
Adult cut size of meal or skipped meal (1 vs. 0)[3]	6.56	5.69	6.55	5.63		
Adult cut size of meal or skipped meal (2 vs. 1)[3]	5.36	6.23	5.21	6.13		
Respondent ate less than felt should	5.88		5.81			
Respondent hungry but did not eat	7.54		7.56			
Respondent lost weight	8.27		8.30			
Adult did not eat for whole day (1 vs. 0)[3]	9.99	8.37	9.97	8.46		
Adult did not eat for whole day (2 vs. 1)[3]	6.98	8.59	7.20	8.71		
Cut size of child's meal	8.99				8.97	
Child skipped meal (1 vs. 0)[3]	11.14	9.15			10.84	9.45
Child skipped meal (2 vs. 1)[3]	7.30	9.29			8.33	9.73
Child hungry, could not afford more food	9.40				9.39	
Child did not eat whole day	11.75				11.51	
Discrimination coefficient	1.00		0.94		1.21	
Number of cases in estimation sample	12,581		24,902		7,814	

[1]The models for the adult and child scales were constrained to have the same mean and standard deviation of Rasch item parameters as the same items in the household scale in order to facilitate comparison across scales. The discrimination coefficients tabled are those required to equate the standard deviations.

[2]The "dichotomous parameters," sometimes referred to as "Rasch-Thurstone parameters," for the two thresholds of the trichotomous items represent the severity level at which response in any level above the threshold is .5. The dichotomous parameters were calculated from the two Rasch parameters for each item; they are similar in practical terms to the parameters for the dichotomous items and are therefore more readily comparable with them.

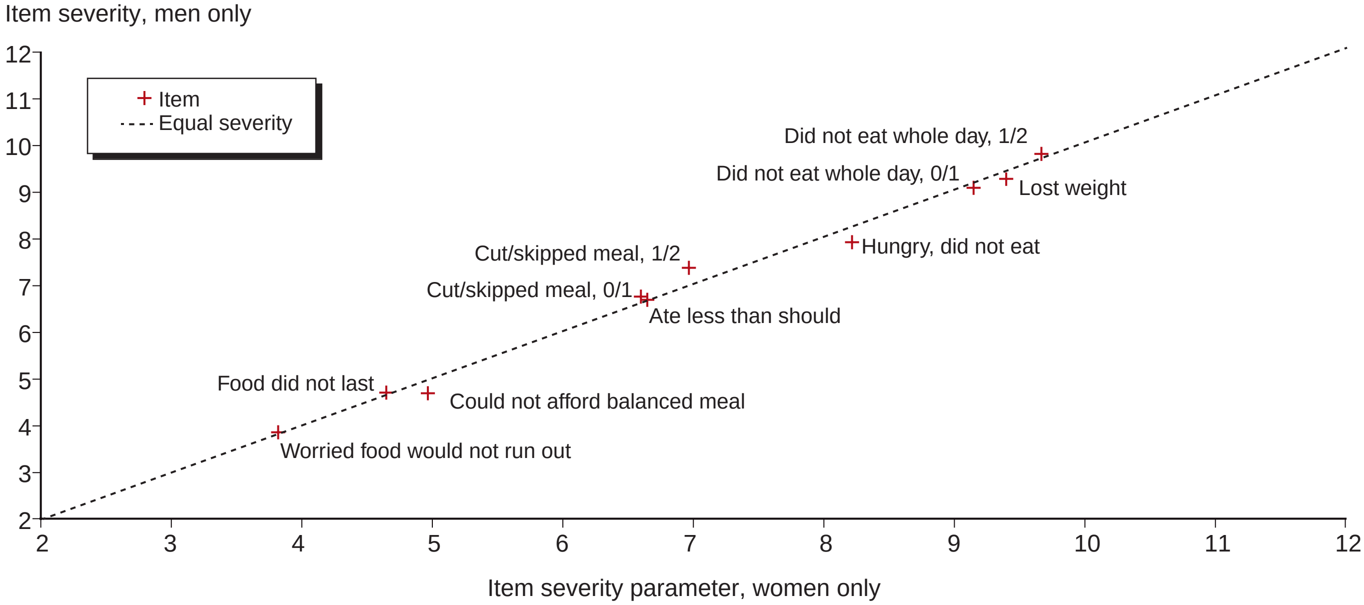
[3]These items were modeled as trichotomies. The parameters for the lower and upper thresholds are tabled.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau Current Population Survey Food Security Supplements for 2003, 2004, 2005 for households with incomes less than 185 percent of the Federal poverty line.

Appendix C—Differential Item Function Between Selected Subpopulations

Figure C-1

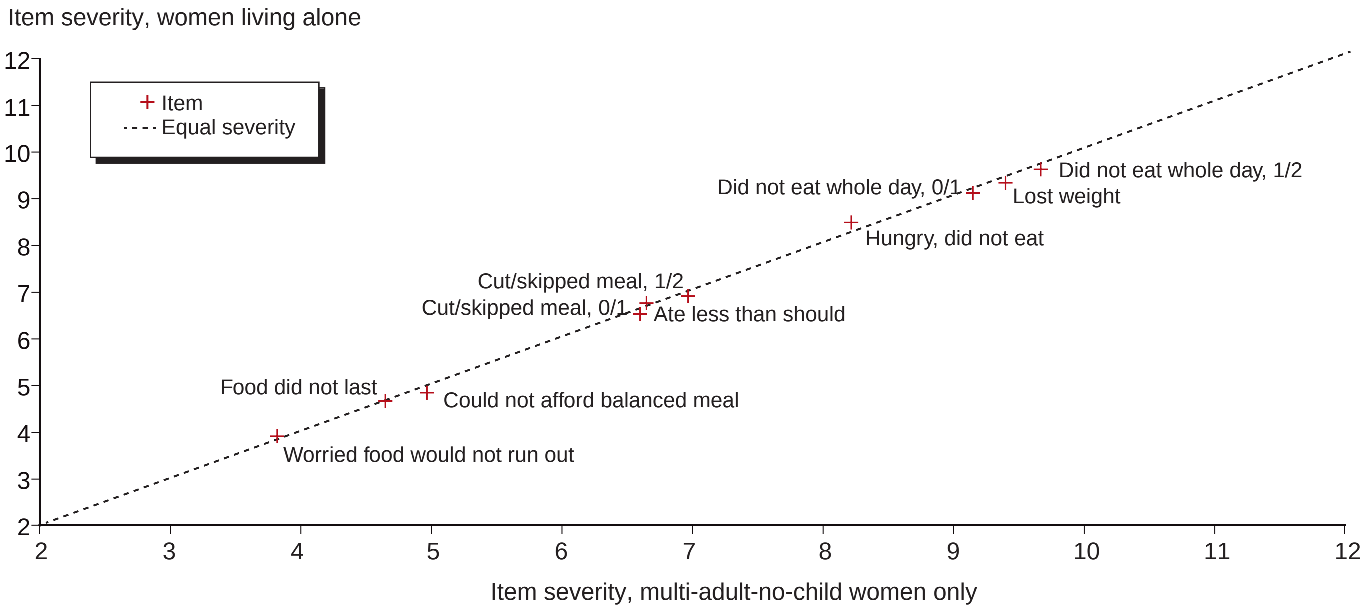
Item-severity comparison, multi-adult-no-child men only versus women only



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007, for households with annual incomes less than 185 percent of the Federal poverty line.

Figure C-2

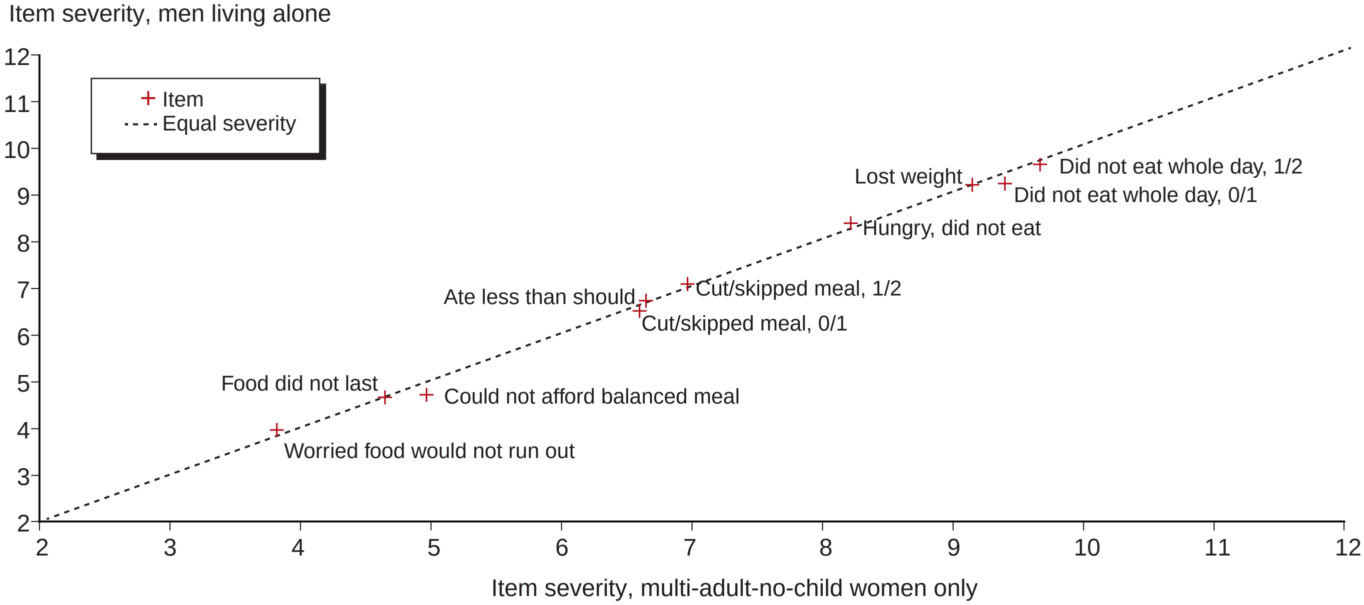
Item-severity comparison, women living alone versus multi-adult-no-child women only households



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007, for households with annual incomes less than 185 percent of the Federal poverty line.

Figure C-3

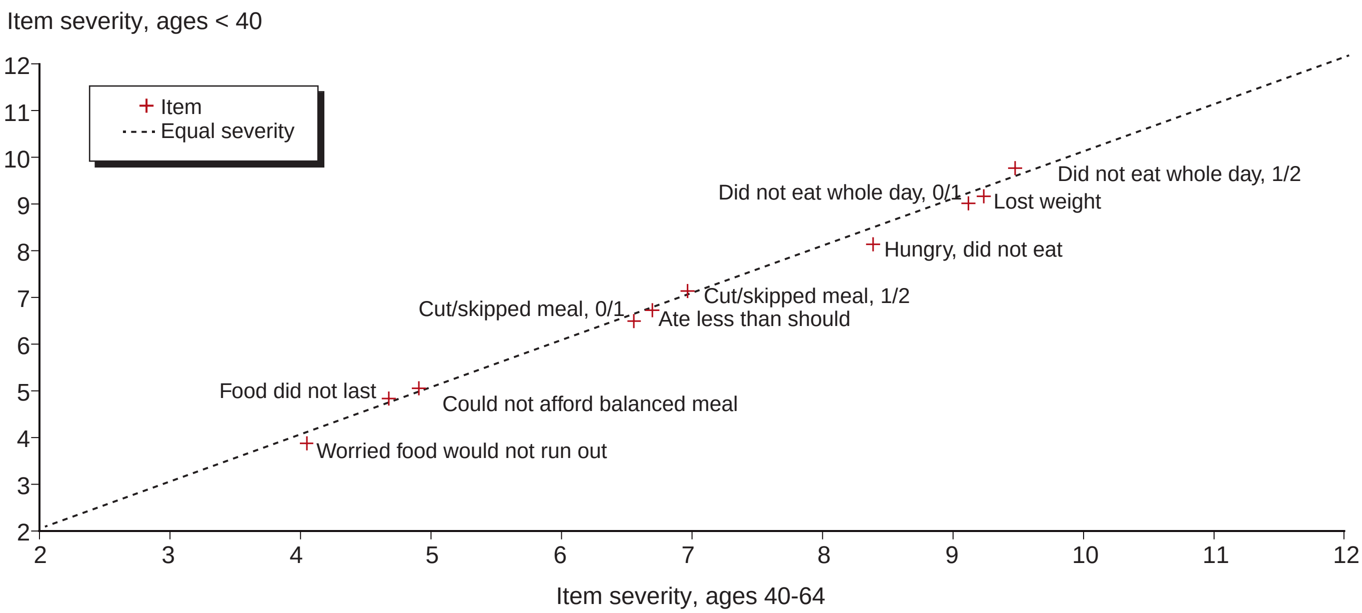
Item-severity comparison, men living alone versus multi-adult-no-child women only households



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007, for households with annual incomes less than 185 percent of the Federal poverty line.

Figure C-4

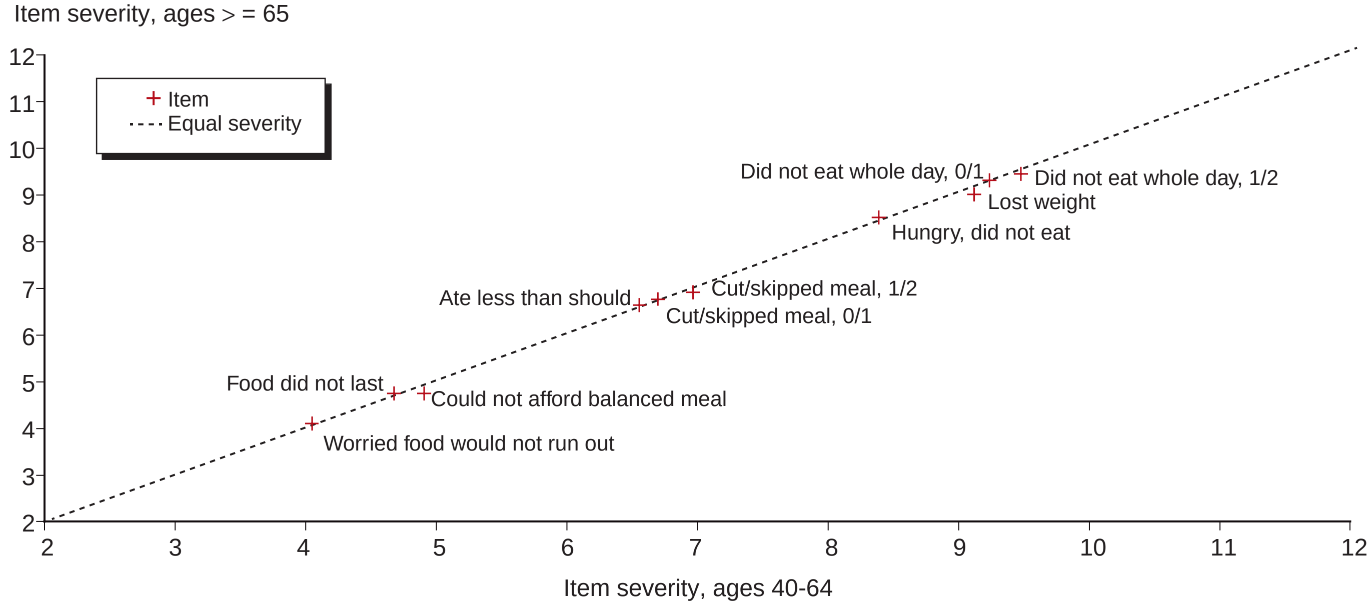
Item-severity comparison, no-child households, all persons age < 40 years versus all persons age 40-64



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007, for households with annual incomes less than 185 percent of the Federal poverty line.

Figure C-5

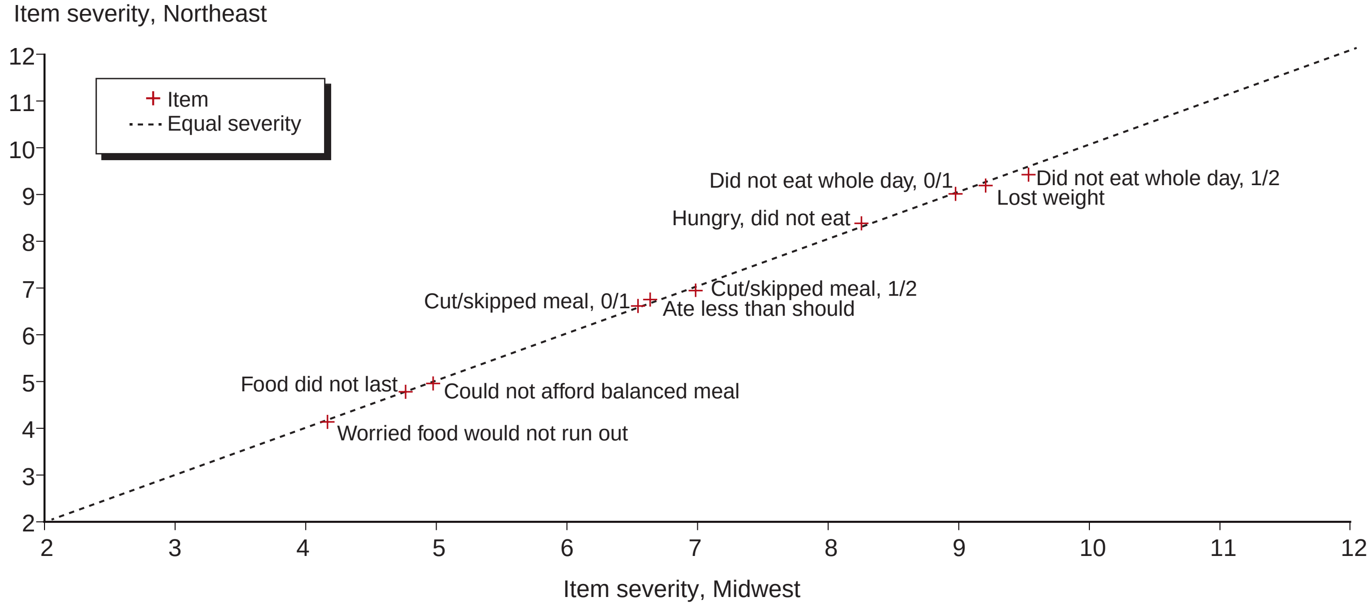
Item-severity comparison, no-child households, all persons age > = 65 versus all persons age 40-64



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007, for households with annual incomes less than 185 percent of the Federal poverty line.

Figure C-6

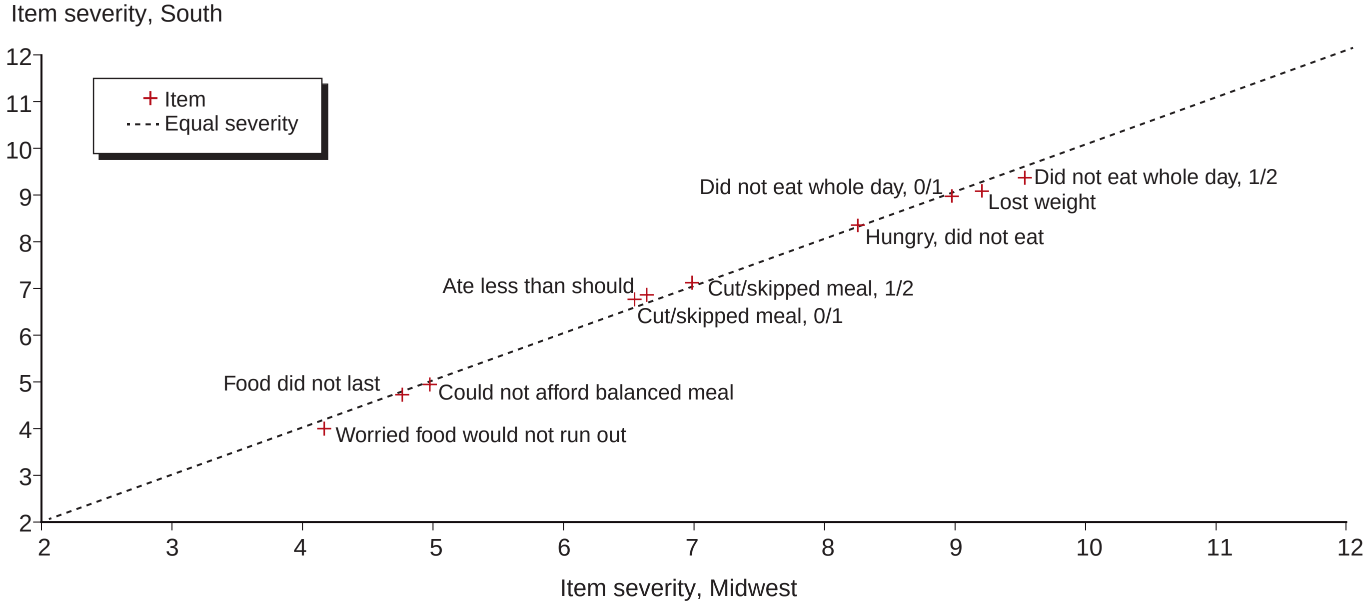
Item-severity comparison, no-child households, Northeast versus Midwest



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007, for households with annual incomes less than 185 percent of the Federal poverty line.

Figure C-7

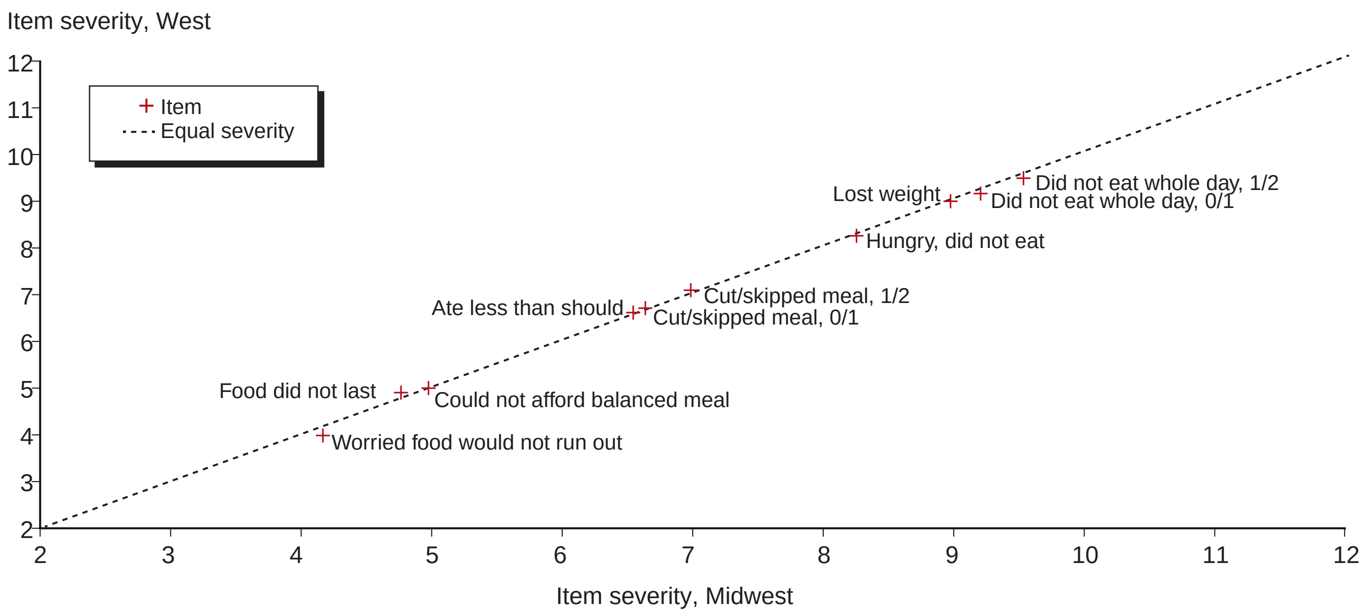
Item-severity comparison, no-child households, South versus Midwest



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007, for households with annual incomes less than 185 percent of the Federal poverty line.

Figure C-8

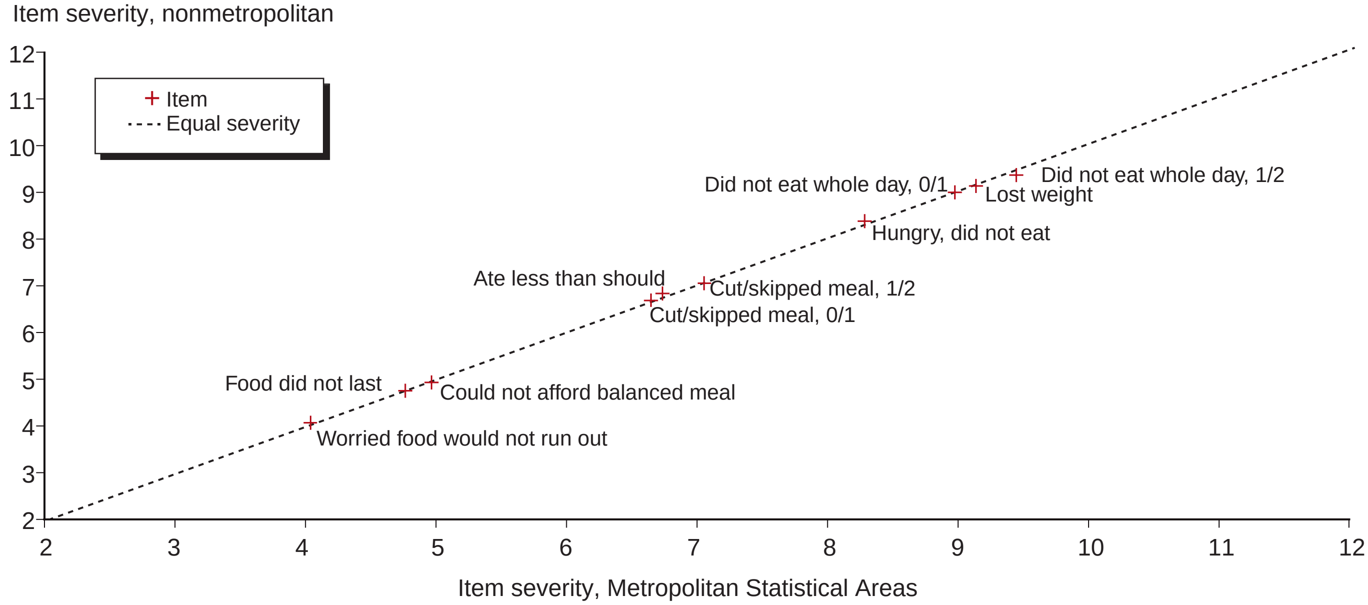
Item-severity comparison, no-child households, West versus Midwest



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007, for households with annual incomes less than 185 percent of the Federal poverty line.

Figure C-9

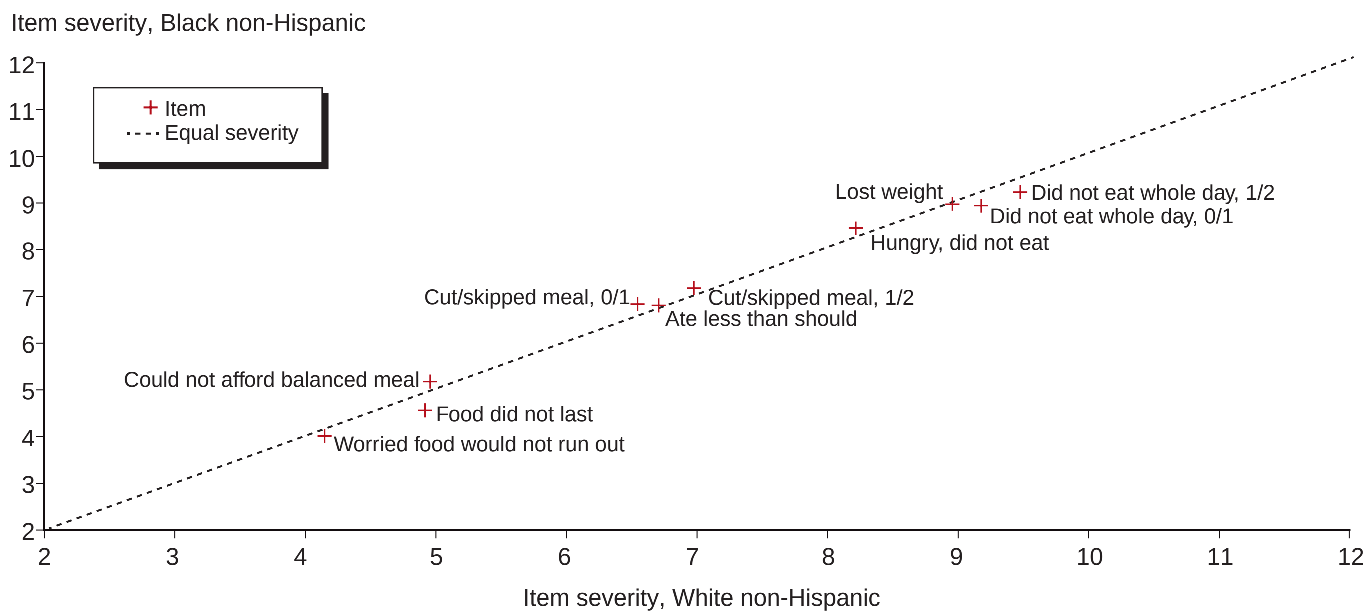
Item-severity comparison, no-child households, nonmetropolitan versus metropolitan



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007, for households with annual incomes less than 185 percent of the Federal poverty line.

Figure C-10

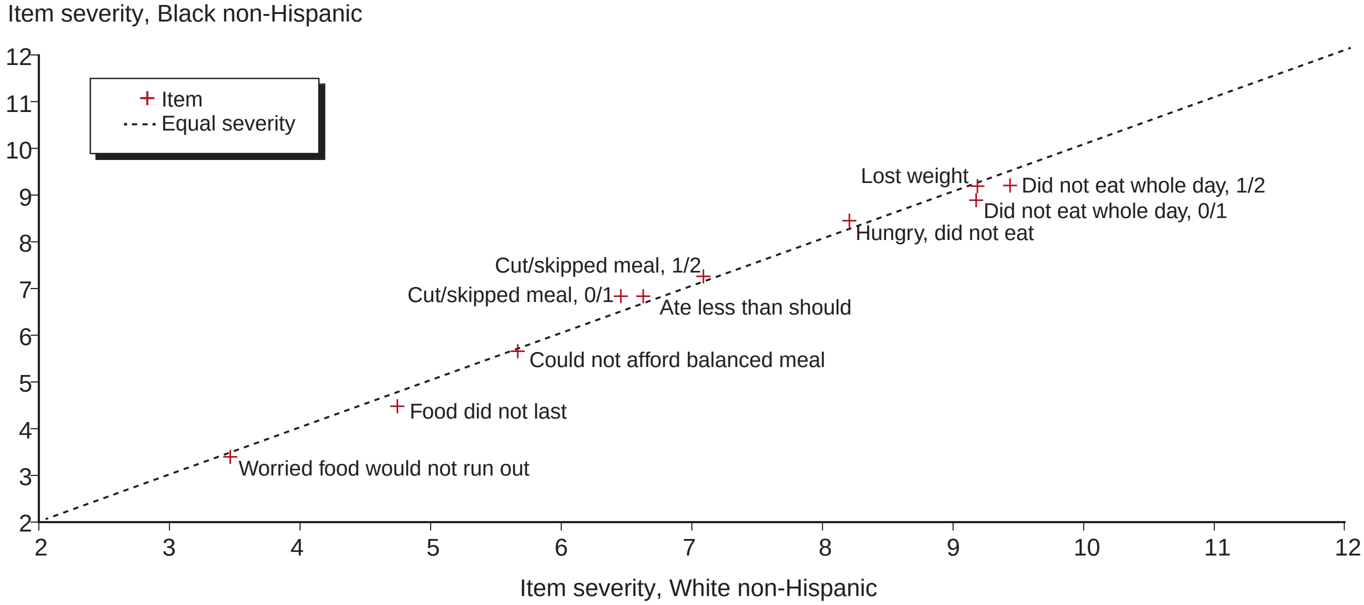
Item-severity comparison, no-child households, Black non-Hispanic versus White non-Hispanic



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007, for households with annual incomes less than 185 percent of the Federal poverty line.

Figure C-11

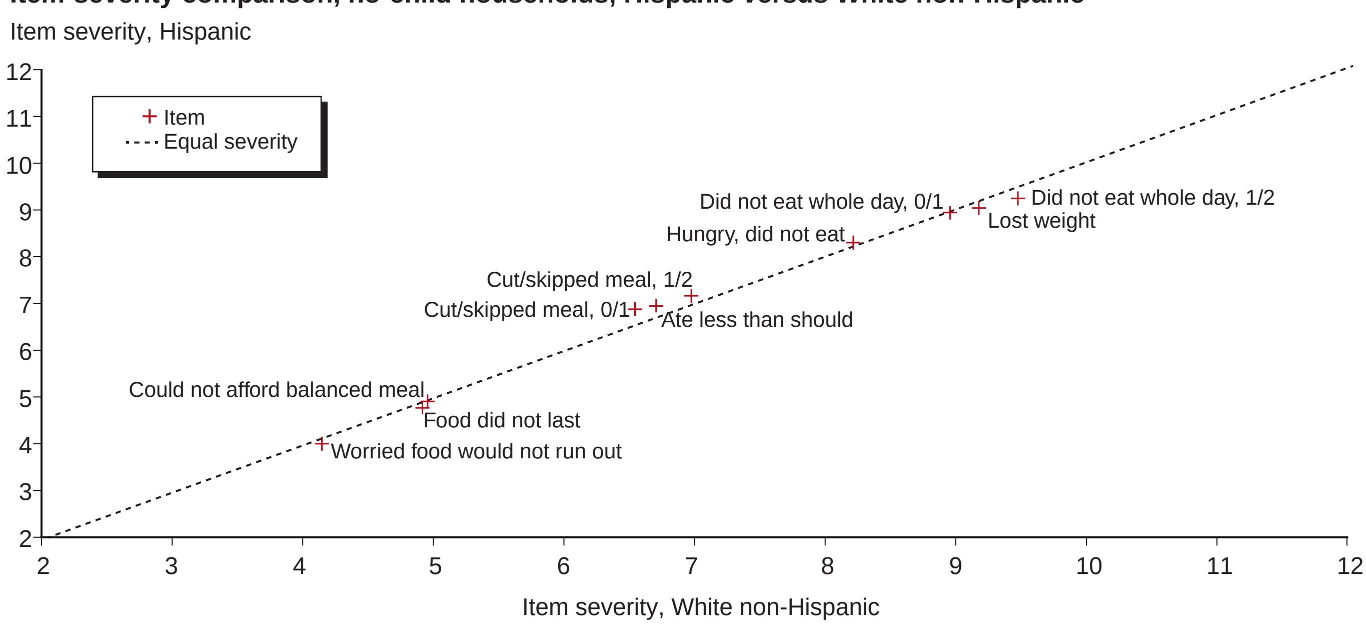
Item-severity comparison, households wtih child(ren), Black non-Hispanic versus White non-Hispanic



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007, for households with annual incomes less than 185 percent of the Federal poverty line.

Figure C-12

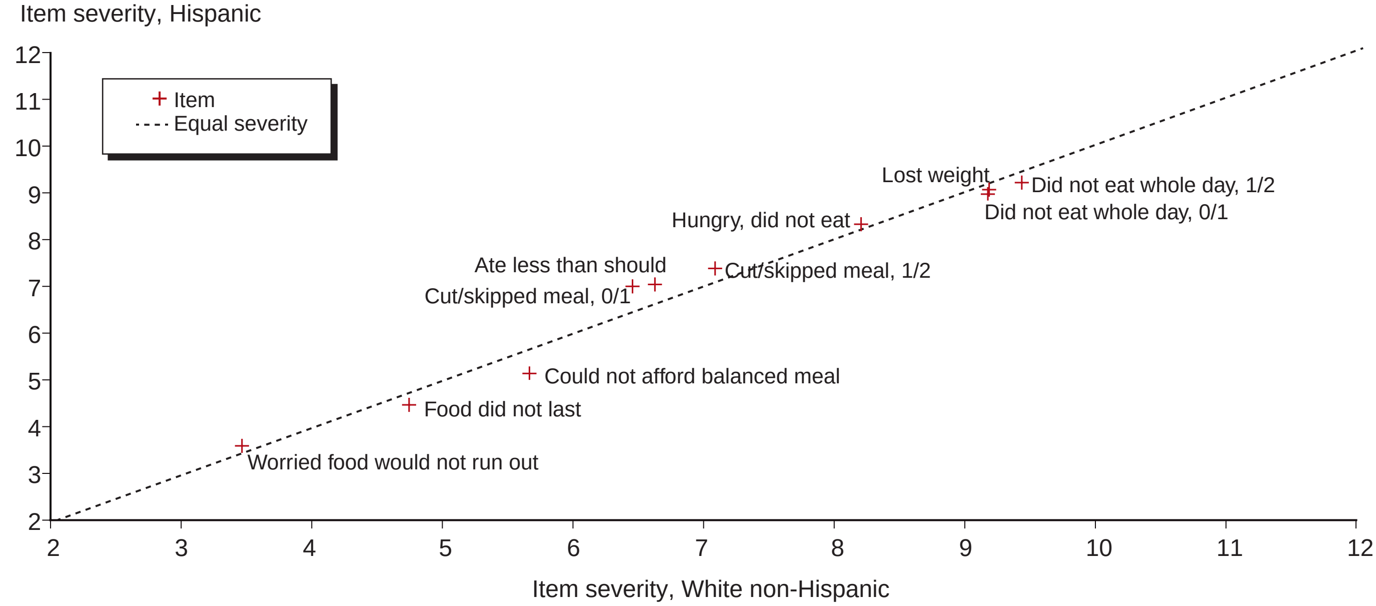
Item-severity comparison, no-child households, Hispanic versus White non-Hispanic



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007, for households with annual incomes less than 185 percent of the Federal poverty line.

Figure C-13

Item-severity comparison, households with child(ren), Hispanic versus White non-Hispanic



Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau, Current Population Survey Food Security Supplements of 1998-2007, for households with annual incomes less than 185 percent of the Federal poverty line.

Appendix D—Item Parameters, Household Parameters, and Raw-Score-Based Probabilities of Food Insecurity and Very Low Food Security for Scales in Chapter 5

Item parameters were estimated for each of the three scales independently using CML methods implemented in ERSRasch. ERSRasch is a suite of SAS programs developed by the author. The CML estimation algorithm treats pairs of dependent items (e.g., *Ever cut or skipped meals* and *Did so in 3 or more months*) appropriately as single trichotomous items. Screens within the food security module are also taken into account by omitting response patterns that cannot occur (such as affirming an item after the first screen while denying all items prior to the screen) from the set of possible responses on which item probabilities for a given raw score are calculated.[1] Screening was not taken into account in modeling the child scale because the adult items, which mainly determine the screens, are not part of the scale.

Prior to calculating household severity parameters, the adult and child scales were adjusted to the metric of the household scale so that the household measures would be directly comparable. Each item parameter was adjusted by a linear transformation so as to equate the mean and standard deviation of the items with those of the same items in the household scale (table D-1). Household parameters and measurement errors were than calculated based on the transformed item parameters, using maximum likelihood methods and taking into account the discrimination parameters implied by the multiplicative constant in the linear transformation (tables D-2, D-3, D-4, D-5). Characteristics of the transformed adult scale were used both for the adult scale for all households and for households with no children in the household scale.

The proportion of households in each raw score that are food insecure—which can also be thought of as the probability that a given household with that raw score is food insecure—was calculated as the proportion of a normal distribution with mean at the household severity parameter and standard deviation equal to the measurement error that would fall beyond a specified threshold. The calculation of the threshold is described in the methods section of the paper. The same procedure was then repeated vis-à-vis a threshold for very low food security.

[1]To reduce respondent burden in the CPS, and to avoid annoying respondents with inappropriate questions, respondents who say “never” to the first three questions (or five questions in households with children) are not administered the remaining food security questions, and are assumed to have high food security. A similar screen is assessed prior to asking about not eating for a whole day. Respondents who have said “no” or “never” to the questions since the first screen are not administered the remaining items, and responses of “no” are assumed in assigning food security.

Table D-1

Severity parameters and infit statistics of items in the household, adult, and child food security scales

Item[1]	Item severity[2] (Rasch-Thurstone threshold)[3] Combined adult-child scale	Adult scale	Child scale	Item infit Combined adult-child scale	Adult scale	Child scale
Worried food would run out before (I/we) got money to buy more	2.60	2.83		1.03	1.00	
Food bought didn't last and (I/we) didn't have money to get more	3.68	3.78		.97	.90	
Couldn't afford to eat balanced meals	4.61	4.34		1.04	1.18	
Adult(s) cut size of meals or skipped meals[3]	6.65 (5.81)	6.58 (6.24)		1.11	.95	
Adult(s) cut size of meals or skipped meals in 3 or more months[3]	5.52 (6.37)	5.36 (6.24)		1.22	1.20	
Respondent ate less than felt he/she should	5.97	5.87		.89	.77	
Respondent hungry but didn't eat because couldn't afford	7.57	7.55		.86	.81	
Respondent lost weight	8.41	8.35		.91	.86	
Adult(s) did not eat for whole day[3]	9.69 (8.24)	9.78 (8.42)		1.04	1.03	
Adult(s) did not eat for whole day in 3 or more months[3]	7.06 (8.51)	7.35 (8.71)		1.19	1.20	
Relied on few kinds of low-cost food to feed child(ren)	3.99		3.66	1.07		1.03
Couldn't feed child(ren) balanced meals	5.58		5.39	.89		.80
Child(ren) were not eating enough	6.97		7.01	.99		.79
Cut size of child(ren)'s meals	8.84		8.87	.96		.96
Child(ren) were hungry	9.23		9.24	.83		.85
Child(ren) skipped meals[3]	10.83 (8.99)		10.49 (9.25)	.97		.86
Child(ren) skipped meals in 3 or more months[3]	7.32 (9.16		8.34 (9.59)	1.16		1.56
Child(ren) did not eat for whole day						
Child(ren) did not eat for whole day	11.46		11.22	.95		.97
Discrimination parameter[2]	1.000	.972	1.263			
Number of cases	35,582	68,564	21,866			

[1]The complete wording of each item includes explicit reference to resource limitation, e.g., "…because (I was/we were) running out of money to buy food," or "…because there wasn't enough money for food."

[2]The three scales were estimated independently. Item parameter estimates for the adult and child scales were then adjusted so as to have the same mean and standard deviation as the same items in the household scale. The difference in standard deviation is reflected in the discrimination parameters.

[3]Each of these pairs of items was estimated as a single trichotomous item. Severity parameter estimates are tabled, and the corresponding Rasch-Thurstone thresholds shown in parentheses.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau Current Population Survey Food Security Supplements for 1998-2006 for households with incomes less than 185 percent of the Federal poverty line.

Table D-2

Measurement characteristics of household food security scale for households with children present

Raw score	Percentage of all households that have this raw score and have children present	Severity parameter for households with this raw score (mean)[1]	Measurement error (standard deviation of true severity of households with this raw score)[1]	Probability of food insecurity for households with this raw score (threshold = 4.251)	Probability of very low food security for households with this raw score (threshold = 6.690)
0[2]	25.981	NA	NA	0.0000	0.0000
1	2.119	2.347	1.171	.0520	.0001
2	1.645	3.415	.937	.1863	.0002
3	1.307	4.198	.840	.4750	.0015
4	.981	4.847	.771	.7801	.0084
5	.872	5.395	.712	.9459	.0345
6	.763	5.873	.676	.9918	.1134
7	.465	6.321	.667	.9990	.2902
8	.386	6.771	.675	.9999	.5478
9	.309	7.228	.673	1.0000	.7880
10	.248	7.664	.645	1.0000	.9346
11	.148	8.057	.610	1.0000	.9875
12	.107	8.414	.589	1.0000	.9983
13	.083	8.758	.588	1.0000	.9998
14	.046	9.115	.613	1.0000	1.0000
15	.038	9.530	.683	1.0000	1.0000
16	.017	10.100	.847	1.0000	1.0000
17	.016	11.133	1.223	1.0000	1.0000
18[3]	.007	NA	NA	1.0000	1.0000
Total	35.539				

NA = parameters cannot be calculated for extreme raw scores.

[1]Household severity parameters and measurement errors are maximum likelihood estimates based on the item parameters for the household scale in table A-1.

[2]Households with raw score zero were assumed to have zero probability of being food insecure; those with raw score 18 were assumed to have very low food security with certainty.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau Current Population Survey Food Security Supplements for 1998-2006 for households with incomes less than 185 percent of the Federal poverty line.

Table D-3

Measurement characteristics of household food security scale applied to households with no children present[1]

Raw score	Percentage of all households that have this raw score and have no child present	Severity parameter for households with this raw score (mean)[2]	Measurement error (standard deviation of true severity of households with this raw score)[2]	Probability of food insecurity for households with this raw score (threshold = 4.251)	Probability of very low food security for households with this raw score (threshold = 6.690)
0[3]	55.012	NA	NA	0.0000	0.0000
1	2.385	2.729	1.254	.1124	.0008
2	1.818	3.968	1.042	.3929	.0045
3	1.683	4.904	.920	.7610	.0261
4	.747	5.638	.830	.9527	.1024
5	.668	6.292	.827	.9932	.3151
6	.756	7.003	.886	.9991	.6382
7	.567	7.771	.873	1.0000	.8922
8	.360	8.474	.844	1.0000	.9828
9	.193	9.263	1.014	1.0000	.9944
10[3]	.272	NA	NA	1.0000	1.0000
Total	64.461				

NA = parameters cannot be calculated for extreme raw scores.

[1]Household severity parameters and measurement errors for this scale are the same as for the adult scale in table A-4, but the thresholds for food insecurity and food insecurity for this scale were constrained to be the same as those for the household scale as applied to households with children. The distribution of households by raw score in this table represents households with the specific raw score and with no child present as a percentage of all households (with and without children).

[2]Household severity parameters and measurement errors are maximum likelihood estimates based on the item parameters for the adult scale in table A-1 and discrimination coefficient 0.972.

[3]Households with raw score zero were assumed to have zero probability of being food insecure; those with raw score 10 were assumed to have very low food security with certainty.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau Current Population Survey Food Security Supplements for 1998-2006 for households with incomes less than 185 percent of the Federal poverty line.

Table D-4

Measurement characteristics of adult food security scale[1]

Raw score	Percentage of all households that have this raw score	Severity parameter for households with this raw score (mean)[2]	Measurement error (standard deviation of true severity of households with this raw score)[2]	Probability of food insecurity for households with this raw score (threshold = 4.573)	Probability of very low food security for households with this raw score (threshold = 6.722)
0[3]	81.455	NA	NA	0.0000	0.0000
1	4.734	2.729	1.254	.0707	.0007
2	3.972	3.968	1.042	.2807	.0041
3	3.548	4.904	.920	.6404	.0240
4	1.455	5.638	.830	.9003	.0956
5	1.279	6.292	.827	.9812	.3015
6	1.347	7.003	.886	.9970	.6245
7	.939	7.771	.873	.9999	.8852
8	.585	8.474	.844	1.0000	.9811
9	.293	9.263	1.014	1.0000	.9939
10[3]	.394	NA	NA	1.0000	1.0000
Total	100.000				

NA = parameters cannot be calculated for extreme raw scores.

[1]Household severity parameters and measurement errors for this scale are the same as for the household scale applied to households with no child present in table A-3, but the thresholds for food insecurity and food insecurity for that scale were constrained to be the same as those for the household scale as applied to households with children. The distribution of households by raw score in this table represents all households (with and without children).

[2]Household severity parameters and measurement errors are maximum likelihood estimates based on the item parameters for the adult scale in table A-1 and discrimination coefficient 0.972.

[3]Households with raw score zero were assumed to have zero probability of being food insecure; those with raw score 10 were assumed to have very low food security with certainty.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau Current Population Survey Food Security Supplements for 1998-2006 for households with incomes less than 185 percent of the Federal poverty line.

Table D-5

Measurement characteristics of children's food security scale[1]

Raw score	Percentage of households with children present that have this raw score	Severity parameter for households with this raw score (mean)[2]	Measurement error (standard deviation of true severity of households with this raw score)[2]	Probability of food insecurity for households with this raw score (threshold = 5.569)	Probability of very low food security for households with this raw score (threshold = 9.424)
0[3]	83.212	NA	NA	0.0000	0.0000
1	8.033	4.43	1.22	.1762	.0000
2	4.512	6.16	1.14	.6995	.0021
3	3.004	7.67	1.02	.9805	.0426
4	.596	8.67	.76	1.0000	.1618
5	.296	9.28	.65	1.0000	.4113
6	.148	9.83	.70	1.0000	.7195
7	.158	10.72	1.02	1.0000	.8980
8[3]	.041	NA	NA	1.0000	1.0000
Total	100.000				

NA = parameters cannot be calculated for extreme raw scores.

[1]Household severity parameters and measurement errors are maximum likelihood estimates based on the item parameters for the child scale in table A-1 and discrimination coefficient 1.263.

[2]Household severity parameters and measurement errors are maximum likelihood estimates based on the item parameters for the adult scale in table A-1 and discrimination coefficient 0.972.

[3]Households with raw score zero were assumed to have zero probability of having food insecurity among children; those with raw score 8 were assumed to have very low food security among children with certainty.

Source: USDA, Economic Research Service analysis of data from U.S. Census Bureau Current Population Survey Food Security Supplements for 1998-2006 for households with incomes less than 185 percent of the Federal poverty line.