Economic Research Service

efan

Electronic Publications from the Food Assistance & Nutrition Research Program

E-FAN-02-010

June 2002

Household Food Security in the United States, 1998 and 1999

Technical Report

By Barbara Cohen, IQ Solutions; Mark Nord, Economic Research Service; and Robert Lerner, James Parry, and Kenneth Yang, IQ Solutions

ERS contact: Mark Nord

Abstract

This report complements prior reports on measuring household food security in the United States. It explores key technical issues related to Current Population Survey Food Security Supplement data, focusing especially on the August 1998 and April 1999 surveys. These technical issues include the estimation of standard errors using either balanced repeated replication techniques or generalized variance functions (GVFs) developed by the Census Bureau; the effect of alternating survey periods between spring and fall for the 1995-99 CPS Supplement; and the effect of using different Item Response Theory (IRT) modeling approaches and software to create the food security scale. The report also presents 1998 and 1999 item calibrations and household scores developed through the use of IRT modeling.

Acknowledgments

We would like to thank Larry Radbill of Mathematica Policy Research, Inc., who provided technical expertise and advice.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Methods used to measure food security and food insecurity with precision have been refined since the initial implementation of the Food Security Supplement to the Current Population Survey (CPS) in 1995. This report provides a technical discussion on several key methodological issues related to the CPS Food Security Supplement data, including techniques used to estimate standard errors, the effects of alternation of survey periods between Spring and Fall, and different item response theory (IRT) modeling approaches used to create the food security scale.

Estimating Standard Errors

Standard errors computed by most common statistical packages are not accurate estimates of the standard errors for the prevalence estimates calculated from a complex-sample survey such as the CPS. Given the limited sample-design information in the publicly available CPS data, the most accurate standard error estimates are produced using balanced repeated replication methods (BRR). However, not all researchers will have the software needed to adequately implement BRR methods. The Census Bureau's generalized variance functions (GVF's), with appropriate adjustments, can also be used to approximate standard errors of food security prevalence rates. The GVF-based standard errors are easier and less costly to calculate than BRR estimates, but they generally underestimate standard errors by approximately 25 percent.

Impact of Survey Season on Food Security Prevalence

The Food Security Supplement has been included in the CPS yearly since 1995, but the data have not been collected in the same month in all years. The months of collection were as follows: April 1995; September 1996; April 1997; August 1998; April 1999 and September 2000. Beginning in 2001, the Food Security Supplement will be fielded annually in early December. An analysis of prevalence rates of food insecurity and hunger across five years was conducted to determine whether the alternation between Spring and Fall data collection introduced a "seasonality" component into the annual estimates. The findings suggest that survey season did affect the measured prevalence of food insecurity and hunger.

Calculating Item Calibrations and Household Scale Scores

The food security scale was developed using statistical methods based on the Rasch measurement model. The model assumes an underlying continuum on which both items and households can be located. Two Rasch modeling approaches have been used to calculate item calibrations, marginal maximum likelihood estimation (MML) and joint or unconditional, maximum likelihood estimation (JML). Item calibrations produced by BILOG's MML procedures are very similar to those produced by JML procedures but are not identical. They differ somewhat because the conditions that the two methods impose on the maximum likelihood solution differ. Further, characteristics of the BILOG software in conjunction with characteristics of the food security data limit the precision of the MML estimates and require special handling. These differences do not threaten the meaning or reliability of the measure.

1. INTRODUCTION

The development of a succinct and accurate method to measure food security with precision began in the early 1990s as researchers developed a set of questions to capture the dimension underlying individual and household indicators (Campbell, 1991). The objective was to assess whether households have access to sufficient quantities of food to fully meet their basic needs – that is, whether they are food secure or insecure (Campbell, 1991; Cohen and Burt, 1990). These concepts were refined by an expert working group of the American Institute of Nutrition and were published by the Life Sciences Research Office (LSRO) of the Federation of American Societies for Experimental Biology (Anderson, 1990). (See *Household Food Security in the United States, 1998 and 1999: Detailed Statistical Report*, Cohen et al, 2002, for a full description of the conceptual and practical development of food security measurement).

Using the April 1995 CPS data, USDA, USDA contractors, and cooperating Federal agencies developed a food security scale based on 18 of the CPS questions. This was done using an Item Response Theory (IRT) statistical model, which posits an underlying latent variable (in the present context, food insecurity and hunger) that cannot be observed directly but can be estimated from respondent answers to a set of instrument items. A relative "severity" is calculated for each of the 18 survey questions on which the food security model is based, ranging from such low-severity items as whether the respondent "worried whether our food would run out" to very severe items, such as a child skipping a meal because no food was available. The household's food security scale score is computed on the basis of the total number of affirmative answers to the 18 increasingly severe food security questions (or 10 questions if no children are present in the household). This scale is then divided into ranges of severity that categorize households as being food secure, food insecure with no hunger evident or food secure with hunger evident.

A series of USDA papers and reports have presented and compared the prevalence of food insecurity and hunger, nationally and by state for 1995-1999, and have begun to explore various technical measurement issues. The reports include: *"Household Food Security in the United States in 1995, Summary Report of the Food Security Measurement Project."* Hamilton, W. et al., 1997a; *"Household Food Security in the United States in 1995, Technical Report."* Hamilton, W. et al., 1997; *"Household Food Security in the United States, 1995-1997: Technical Issues and Statistical Report."* Ohls, J., L. Radbill, and A. Schirm, 2001; *"Household Food Security in the United States, 1995-1998, Advance Report."* Bickel, G., S. Carlson, and M. Nord, 2000; *"Prevalence of Food Insecurity and Hunger, by State, 1996-1998"* Nord, M., K. Jemison, and G. Bickel, 1999; *"Household Food Security in the United States, 1999."* Andrews, M., M. Nord, G. Bickel and S. Carlson, 2000; *"Household Food Security in the United States, 1998 and 1999: Detailed Statistical Report*." Cohen, B., J. Parry, and K. Yang, 2002.

This report complements these reports by analyzing and discussing four key technical issues of the CPS Food Security Supplement. These include: the development of procedures for estimating standard errors; an assessment of the feasibility of using the generalized variance functions (GVFs) developed by the Census Bureau for estimating standard errors, instead of directly estimating standard errors using balanced repeated replication (BRR) procedures; the effect of the alternation of survey periods between April and August/September for the CPS Supplement over the years 1995 - 1999; and the implications of using different IRT modeling approaches and software to create the food security scale. Prior to presenting these issues in chapters 3-5, the following chapter presents data issues that are unique to the 1998 and 1999 Food Security Supplements.

2. UNIQUE DATA ISSUES IN THE 1998 AND 1999 FOOD SECURITY SUPPLEMENTS

Although the contents of the Food Security Supplements for 1998 and 1999 are essentially unchanged from those in 1995, 1996 and 1997, the supplement was substantially reorganized in 1998. The main series of questions was reordered to approximate the severity order of the items (as established by statistical analysis of data from the 1995 survey) and the corresponding variables were renamed to reflect the new questionnaire structure. The reordering allowed insertion of two internal screens and a less stringent initial screen that are described below. The major content areas of the Food Security Supplement questionnaire are:

- [Labor Force Survey – precedes supplement]
- Food Expenditures
- Preliminary Screener
- Use of Federal Food and Nutrition Assistance Programs
- Food Sufficiency Screener
- Food Security
    - Block 1
    - "Internal" Screen
    - Block 2
    - "Internal" Screen
    - Block 3
- Ways of Avoiding or Ameliorating Food Deprivation
- Minimum Needed Food Spending

Other changes include: (1 – in 1998 only) a single "usual" household food expenditure question which replaced the series of items on actual food spending; (2) two split ballot sets of experimental questions (described below); (3) a more complete set of "how often did this occur?" follow-up questions to the main food security and hunger series; and (4) addition of a final question that asks the respondent what would be the lowest amount their household could spend for food per week of per month and still provide a healthy acceptable diet. Also, child referenced questions in households with only one child were referenced to "your child." In previous years, these questions were referenced to the child by name.

A. Screening of the Food Security Supplement

The Food Security Supplement includes several screens to reduce respondent burden and to avoid embarrassing respondents by asking them questions that are inappropriate given other information they have provided in the survey. Some of the screener variables use information from the monthly labor force core data as well as from the Food Security Supplement. In 1998 and 1999, households with income above 185 percent of the poverty threshold[1] and who responded "no" to HES2 were skipped over the questions on participation in food assistance programs. Households with income above 185 percent of poverty who registered no indication of food stress on preliminary screener questions (HES2, HESS1, or HESS1A) were skipped over the rest of the "Food Sufficiency and Food Security" section and the "Ways of Avoiding or Ameliorating Food Deprivation" section. There are also two "internal" screeners in the main food security section (the questions that are used to calculate the household food security scale). This series of questions is divided into three blocks. After the first and second blocks, households that have registered no indication of food stress in the preceding block are skipped over the rest of the "Food Security" section.

The screening rules that determine whether a household was asked the questions used to calculate the food security scale have varied somewhat during the first four years of fielding the Food Security Supplement. These different screening procedures biased estimated prevalence of food insecurity and hunger differently in each year. Adjustments must be made for these differences to compare prevalence of food security and hunger across years.

Screeners also were applied based on whether the household included any children, so that households without children were not asked questions which refer specifically to children. This screener, as calculated at the time of the survey, classified as children all persons age 17 or younger. However, for processing and analyzing the

[1] Using interview data on total household income (HUFAMINC) and the number of people in the household (HRNUMHOU), along with the 1998 and 1999 poverty guidelines, a variable (HRPOOR) was developed to define households as being above or below the 185% poverty threshold.

food security data, household reference persons or spouses of household reference persons (PERRP=1, 2, or 3) are not considered children even if they are age 17 or younger. Therefore, child-referenced questions are recoded to “missing” in households in which the only person age 17 or younger is a reference person or spouse of a reference person. The food security scale, status, and screener variables reflect this recoding, however the original responses to each item are not recoded in the public-use data file, and the user will need to recode these if they are to be analyzed or used to replicate scale scores.

B. Experimental Questions

There are two sets of experimental questions that were asked of respondents in only one month-in-sample group:

(1) Households in rotation 4 (HRMIS=4) were asked an experimental variant of the food sufficiency question, HESS1A, instead of HESS1.

(2) Households in rotation 8 (HRMIS=8) were asked several food security questions referenced to the respondent or to a specific child in place of corresponding questions in other month-in-sample groups which referred either to all adults or all children in the household. Adult items that are normally asked of “you or other adults in the household” in multiple-adult households were referenced only to the respondent. Selected items that are normally asked of "the children" in multiple-child households were asked of a specific focus child in these households. The selection of the focus child was randomized with respect to characteristics of interest based on which child's birthday was nearest to the date of interview. As a lead-in to the first such question, the respondent was advised, "The next questions ask about a particular child living in the household; that is (CHILD'S NAME)." In subsequent questions, the child's name was inserted as a referent. Because these questions refer to specific individuals, and not to the experience of all members of the household, it is not possible to calculate scale scores for these households that are precisely comparable with those of other households. For this reason, these households are assigned missing values on food security scale and status variables, and an adjusted set of weights is provided to

account for their exclusion. The focus child in households in rotation 8 is identified by the variable PRSCHILD.

The food security status of households in rotation 8 with more than one adult or more than one child cannot be determined in ways that are comparable with those of other households because of the experimental, individually referenced, questions administered to those households (described above). Adjusted weights, HHFSWGT and PWFSWGT, are provided in the public-use data file for estimating food security and hunger prevalence and for analyses which include the food security scales or food security status variables. For households with one adult and not more than one child, these food security status weights are identical to their supplement weight counterparts. For households with more than one adult or more than one child, the food security status weights are zero in rotation 8 and adjusted by a factor of approximately 8/7 for households in rotations 1-7, so as to represent the same total population and number of households as the core weights and supplement weights do. This is a ratio adjustment, however, not an iterative adjustment and therefore controls for subpopulations or State populations may not exactly match weighted tables.

C. Additional Changes in the 1999 Food Security Supplement

While the 1999 Food Security Supplement was almost identical to that of 1998, there were two changes from 1998:

(1) A series of questions on food spending at various kinds of places plus a follow-up asking about usual spending for food replaced the single "usual" household food expenditure question asked in 1998 and in rotations 1-7 of 1997. (The 1999 series was similar to that used in 1996, and in rotation 8 in 1997.)

(2) A split ballot test of two forms of the USDA/NHANES food sufficiency questions (HESS1 and HESS1A/HESS1A1) replaced a different test carried out in 1998.

The split-ballot test of individually referenced Food Security questions in rotation 8 (HRMIS=8) was continued as in 1998. In 1999, the first and second screeners were administered incorrectly. The food sufficiency questions (HESS1, HESS1A and HESS1A1) were switched late in the process of finalizing the questionnaire, and the screeners did not correctly register the change. In the analysis of 1999 data, editing was used to correct for those who were included but should have been screened out. However, data for those households that were erroneously screened out were lost to the analysis. Since these households should not have been screened out, they could either be considered as being screened in and having all missing data values, or they could remain screened out. After an analysis using data from earlier years showed that the bias on prevalence estimates caused by the screening errors was smaller than the opposite bias that would be caused by excluding the households from analysis, and was not large enough to be worrisome, it was decided that these households should remain screened out.

3. ESTIMATING STANDARD ERRORS

A key activity in the development of prevalence estimates is the determination of their precision. This is important in assessing the overall accuracy of the estimates, and is required in order to determine whether observed changes over time or differences between population subgroups are statistically significant.

Standard errors computed by most common statistical packages are not accurate estimates of the standard errors for the prevalence estimates calculated from a complex-sample survey such as the CPS. The formulae used by programs such as SAS and SPSS are valid only for statistics computed from simple random samples. The CPS, however, does not use a simple random sample. Rather, it is based on a stratified cluster design where addresses from different areas are sampled with different probabilities. Because of unequal initial sampling probabilities together with adjustments needed to account for household non-response, sample weights must be used to estimate population characteristics from the CPS. SAS and SPSS do, if used correctly, deal appropriately with sample weights when computing standard errors. However, these packages do not currently deal with the stratification or clustering in the sample design. Other statistical packages such as STATA and SUDAAN can calculate unbiased standard errors based on complex samples, but CPS does not disclose the sampling information needed to implement these packages.

With the CPS, the combined effects of stratification and clustering produce samples that are less statistically efficient than simple random samples because households from the same cluster tend to be more similar than households randomly drawn from the population, the effective sample size is smaller than the nominal sample size. For that reason, the standard error estimates that fail to account for the CPS sample design (such as those currently produced by SAS or SPSS) will generally be too small.

There are a number of approaches to estimating standard errors of population estimates from a complex sample design such as the CPS. Given the limited sample-design information publicly available on the CPS data, the most accurate standard error estimates are produced using balanced repeated replication methods (BRR). Using the method as implemented in WestVars Complex Samples, version 3.0 (a statistical package implemented as an SPSS supplementary module), standard errors were estimated for publication with the prevalence estimates in *Household Food Security in the United States, 1998 and 1999: Detailed Statistical Report* (Cohen, et al., 2002). Details of the application of these procedures to the 1998 and 1999 food security data are described below. However, not all researchers will have the software needed to adequately implement BRR methods. In the latter part of this chapter we assess the feasibility of using the Census Bureau's generalized variance functions (GVF's), with appropriate adjustments, to approximate standard errors of food security prevalence rates.

A. Balanced Repeated Replication Methods (BRR)

The basic premise of the Balanced Repeated Replication (BRR) methods used to compute standard errors for the food insecurity prevalence estimates from the CPS food security supplements is that data collected is treated as a population, which is subsampled in the same way that the original sample was selected from the larger universe. The estimated statistic is calculated for the total sample and then the total sample is repeatedly subsampled in a way that reflects the sampling design of the total set of data. After the subsampling takes place, the statistic of interest is calculated for each subsample, and the variability among these subsamples is used to estimate the sampling error of the statistic.

More specifically, balanced repeated replication (BRR) methods are generally used with multistage stratified sample designs. After grouping all primary sampling units (PSU's) into strata, two PSU's are selected from each stratum using sampling with replacement. This provides two independent estimates for each stratum. Next a series of "replicate samples" is drawn. In each replicate sample, one of the two psus in each strata is included. A total of 2^n different subsamples is possible (where n is the number of strata). It is not necessary to form all possible replicates because the variance can be estimated with full information using fewer than the full complement of replicates. The minimum number of replicates needed to have full information is the smallest integer that is divisible by 4 and greater than or equal to n.

The calculation of standard errors for the food security prevalence estimate used "State" as the strata and the month in which the household entered the sample (the rotation group coded as HRMIS in the dataset) as the PSU[2]. The variable HRMIS was recoded into two groups: odd month (1,3,5,7) and even (2,4,6,8), which served as the two independent samples within each stratum. The food security weights (person weights for the person files and household weights for the household files) were used as sample weights. There were 52 balanced replicates used to develop these standard errors.

When population totals are known, the precision of the variance estimates can be improved using a poststratification weighting procedure after initial replicate estimates are calculated. This creates adjusted weights for respondents in each replicate so that the sums of the adjusted replicate weights are equal to the known population totals. For this analysis of the Food Security Supplement data in which "States" were used as strata, replicate weights in the two independent samples within each stratum were adjusted to sum to that State's known population total. Within each PSU-stratum combination, all weights were adjusted by the same ratio.

It should be noted, that although the 8th rotation was used to test new versions of the food security supplement with households in which there was more than one adult or more than one child, the cases in this rotation that were the same as those in all other rotations, were included in the calculation of the prevalence estimates and thus were included in the calculation of the standard errors. Households with more than one adult or more than one child were assigned a weight equal to "missing" in the editing of the data file. Thus, their values would automatically be dropped in the calculation of standard errors. Comparisons of prevalence rates between those remaining in the 8th rotation and those in other rotations indicate that this smaller, selective population in the 8th rotation does not create a noticeable bias in the calculation of overall standard errors for the population and the subgroups.

---

[2] Stratifying by state does not completely replicate the original sampling procedure, which is also stratified within states. Therefore, standard errors estimated using this procedure may be somewhat biased downward.

B. Assessment of the Feasibility of Using Generalized Variance Estimates

In addition to directly estimating the relevant standard errors and confidence intervals, we conducted an assessment of the feasibility of using the Census Bureau's generalized variance functions (GVFs) to approximate standard errors of food security prevalence rates. Descriptions of the CPS sample design, weights, procedures for implementing GVFs and relevant tables of factors (often called a and b parameters) are provided by the Census Bureau in Source and Accuracy Statements that accompany their public use data files. The GVFs provided by the Census Bureau are designed to compute standard errors for two types of estimates: subpopulation total numbers and percentages of the population or subpopulations, with a specific characteristic (such as food insecurity). Although the Census Bureau provides instructions for their application to tables with two or more dimensions as well as for their use when computing quantiles (such as medians) and averages, the Census Bureau's GVFs are designed to be used primarily for unidimensional tabulations of population characteristics.

The standard error of an estimated number using the GVFs with food security data is obtained as follows:

$$s_x = \sqrt{ax^2 + bx}$$

S_x is the approximate standard error with x being the size of the estimate (weighted) and a and b are the parameters associated with the particular type of characteristic. For food security measurement purposes, the following a and b parameters are provided by the Census Bureau for the 1998 and 1999 monthly files for household estimation:

	a parameter	b parameter
Total or White Population	-0.000010	2,068
Black Population	-0.000075	1,871
Hispanic Population	-0.000145	3,153

The approximate standard error of an estimated percentage is derived using the following formula:

$$s_{x,p} = \sqrt{\frac{b}{x} p(100 - p)}$$

Here $s_{x,p}$ is the approximated standard error with x being the total (weighted) number of people, or households in the base of the percentage, p is the percentage, and b is the parameter or factor associated with the characteristic in the numerator of the percentage. The b parameter is the same presented above.

The GVF parameters provided by the Census Bureau are based on the sample size of the monthly labor force survey. However, a small proportion of households that complete the labor force survey decline to complete the supplement. Further, in 1998 and 1999, part of the 8^{th} rotation was not used to estimate prevalence rates. Thus, for the purposes of calculating GVF-based standard errors for prevalence estimates calculated from food security supplement data it is necessary to adjust the parameters to reflect the true number of cases in the supplement, accounting for the loss in sample size. To accomplish this, the b parameters above are adjusted as follows where b_a is the adjusted b parameter, N is the total unweighted number of household records in the full CPS and n_s is the unweighted number of cases used to estimate prevalence in the supplement[3]:

$$b_a = b\left(\frac{N}{n_s}\right)$$

Using data for all households in 1998 we estimated standard errors following the GVF procedures with the adjustment for reduced sample size in the supplement (see Table 3.1). The estimated standard errors derived from the GVF procedures were generally smaller than those derived using BRR methods. For the prevalence of food insecurity for all households the GVF-based standard error estimate is .05 percentage points smaller than BRR-based estimate. Among the subpopulations there are four groups for which the GVF-based estimates of standard errors for the prevalence of food insecurity are larger than the BRR estimates: other non-Hispanic households (.21 percentage points), households living with incomes 185 percent of poverty or more (.01

percentage points), households living in the Northeast (.01 percentage points) and those living in the West (.06 percentage points). For most subpopulations, the GVF-based estimates are smaller than the BRR-based estimates with a range between .02 and .8 percentage points. There is no difference in the estimates of food insecure households for elderly people living alone.

A similar pattern is found for estimates of the prevalence of food insecurity with hunger. For all households, the GVF-based estimate is .03 percentage points smaller than the BRR-based estimate. For all but four subpopulations, the GVF-based estimates are smaller than the BRR-based estimates; the range of the difference is between .01 and .23 percentage points. There is one subgroup for which there is no difference between the estimates (those with no children under 18 years of age) and three subgroups for which the GVF-based estimates are larger than the BRR-based estimates: "other" households with children (.05 percentage points), other non-Hispanic households (.14 percentage points) and households living in the West (.07 percentage points).

In general, then, GVF methods appear to underestimate standard errors, which should be kept in mind if they are used with food security statistics. However, both GVF and BRR methods have advantages and drawbacks. One disadvantage of using replication methods is the potential for flaws in the estimates if the design of the replicates does not mimic the full-sample design. These variance estimates will be subject to bias. However, the major disadvantage of replication is the intensive computer effort required (with respect to equipment and person hours) and the need for acquiring appropriate software.

Similarly, there are several drawbacks to using the GVFs. Although these are called *generalized* variance functions, the a and b parameters used when applying the GVFs are specific to (1) the data being used, (2) the weights being used, (3) the outcome (trait) being measured, and (4) the population (group or sub-group) being assessed. For example, in the current application, parameters would be needed for the food security

---

[3] The same ratio (N/n_s), based on the full CPS, was used to adjust the b parameter for all subpopulations.

supplement sample (distinct from the CPS core sample), the food security supplement weights, and food insecurity (the outcome). Working with that combination of sample-weight-outcome, separate sets of parameters would be needed for the full population and for each of the subgroups of interest (such as race, ethnicity, household types, states, regions, or age groups). As an alternative, parameters can sometimes be borrowed from other applications. For example, GVF parameters used for unemployment (from the CPS core) or for poverty (from the March supplement) might be appropriate. Or parameters from one population subgroup can sometimes be applied to other groups (this is the procedure described by the Census Bureau when applying its parameters to tables with two or more dimensions). However, there is no assurance of the accuracy of the standard errors computed in this way.

There are two advantages to using GVF methods. The first is that approximate standard errors for estimates derived from microdata can be calculate without specialized software. The second is that standard errors can be calculated for published statistics.

In summary, if the computer equipment, software and staff time is available, the replication methods will provide more accurate estimates of standard errors. However, the GVF-based estimates are much easier and less costly to calculate and can be substituted for BRR estimates, acknowledging that they generally underestimate standard errors by approximately 25 percent.

TABLE 3.1: Household Food Security Status by Selected Household Characteristics, 1998.				
	STANDARD ERRORS			
	GVF-BASED ESTIMATE		BRR ESTIMATE	
Characteristic	**Food Secure**	**Food Insecure with Hunger**	**Food Secure**	**Food Insecure with Hunger**
	Percentage Points	**Percentage Points**	**Percentage Points**	**Percentage Points**
All Households	0.15	0.09	0.20	0.12
Household Composition				
With Children < 6	0.46	0.23	0.56	0.29
With Children < 18	0.30	0.16	0.38	0.22
Married Couple Families	0.30	0.14	0.43	0.15
Female Head, No Spouse	0.78	0.50	0.88	0.57
Male Head, No Spouse	1.32	0.76	2.12	0.99
Other Households with Child[e]	2.12	1.35	2.17	1.30
With No Children < 18	0.17	0.11	0.20	0.11
More Than One Adult	0.19	0.12	0.27	0.16
Women Living Alone	0.38	0.25	0.40	0.30
Men Living Alone	0.48	0.33	0.51	0.36
Households With Elderly	0.24	0.13	0.29	0.18
Elderly Living Alone	0.37	0.22	0.37	0.26
Race and Hispanic Ethnicity				
White, Non-Hispanic	0.15	0.09	0.18	0.11
Black, Non-Hispanic	0.56	0.37	0.68	0.39
Hispanic[f]	0.88	0.51	1.05	0.55
Other Non-Hispanic	0.84	0.46	0.63	0.32
Household Income-to-Poverty Ratio				
Under 1.00	0.67	0.48	0.72	0.51
Under 1.30	0.55	0.37	0.57	0.39
Under 1.85	0.41	0.26	0.47	0.33
1.85 and Over	0.14	0.07	0.13	0.08
Income Unknown	0.37	0.22	0.76	0.37
Area of Residence[g]				
Inside Metropolitan Area	0.17	0.10	0.23	0.13
In Central City	0.35	0.22	0.37	0.26
Not In Central City	0.22	0.12	0.37	0.18
Outside Metropolitan Area	0.35	0.20	0.39	0.28
Census Geographic Region				
Northeast	0.34	0.20	0.33	0.26
Midwest	0.29	0.16	0.50	0.19
South	0.27	0.16	0.34	0.28
West	0.36	0.21	0.30	0.14

Source: Tabulations of Current Population Survey, Food Security Supplement data.
e,f,g, -- See End Notes

4. IMPACT OF SURVEY SEASON ON FOOD SECURITY PREVALENCE

Although the Food Security Supplement has been included in the Current Population Survey (CPS) yearly since 1995, the data have not been collected in the same month in all years. The months of collection were as follows: April 1995; September 1996; April 1997; August 1998; April 1999 and September 2000. Starting in 2001, the Food Security Supplement will be fielded annually in early December. Prior to that an extra collection was conducted in April 2001. To date, several comparisons have been made across the five-year time period since the supplement was included in the CPS. The most recent report (Andrews, et al, 2000) raised the issue as to whether annual estimates include a "seasonality" component because of the alternation between Spring and Fall collection.

Theoretically this should not be true because people are asked about their experiences for the past 12-month period. However, it is possible that people's answers reflect their most recent experiences more than they do the experiences of 11 or 12 months earlier. In this case, estimates will have a seasonal component if food deprivation varies seasonally in similar ways in a majority of food insecure and near food insecure households. An analysis of food insecurity and hunger prevalence rates across five years, taking into account the season in which data were collected, suggests that survey season does affect the measured prevalence of food insecurity and hunger.

A. Data Sources

The analysis of the impact of the survey season on food security status utilized annual household food insecurity and hunger prevalence statistics. Since there were screening differences in the survey administration across the five-year span, data were adjusted to a common screen for comparability. The analyses used data for all households, focusing on overall food insecurity and food insecurity with hunger. The prevalence rates used for the year 1995 were published in the *"Measuring Food Security in the United States-Household Food Security In The United States 1995-1998 (Advanced*

Report)," (Bickel et al.1999). The prevalence rates for 1996 through 1999 were derived directly from public use data sets.

The analyses were conducted using linear regression models with "year" as the unit of observation. The dependent variable in the first model is the proportion of households that were food insecure. In the second model the dependent variable is the proportion of households that were food insecure with hunger. Independent variables in both models included season (Spring or Fall) and time. Time is measured as the number of months since the first Food Security Supplement was collected (April 1995). The inclusion of a time variable controls for long-term trends in prevalence rates.

B. Results

Tables 4.1 and 4.2 present the results for the estimated regression equations for the effect of season and time on food insecurity and food insecurity with hunger. The season coefficient is relatively large when compared to the standard errors for annual estimates (available for 1998 and 1999). Despite the fact that there are only two degrees of freedom for the season coefficient, it approaches statistical significance ($p=.13$) for food insecurity and is statistically significant for hunger ($p=.04$).

C. Summary

The results above suggest that seasonality has an impact on the annual estimates of food insecurity and food insecurity with hunger. They suggest that food insecurity estimates are 1.15 percentage points higher in the Fall than in the Spring and that food insecurity with hunger estimates are 0.6 percentage points higher in the Fall than in the Spring. However, the residuals suggest that other differences, perhaps related to factors specific to the year of the survey, may also impact the results. It is important to note that the small number of data points limits the reliability of these results. They are consistent with an effect of seasonality but do not provide conclusive evidence. Further exploration using the September 2000 and planned April 2001 surveys will be needed to verify whether season does, in fact, bias prevalence estimates.

Table 4.1: Effect of Survey Season on Measured Prevalence of Food Insecurity, Regression Analysis Results

Number of observations	=	5
R-squared	=	.829
Coefficients and Significance:		
Season = Fall	=	1.150
Significance	=	P=.13
Time (months)	=	-.0316
Significance	=	P=.16

Case	Season	Percentage of Households with Food Insecurity	Percentage of Households with Food Insecurity (Standard Error)	Residual
1995	Spring	10.3	n.a.	.33
1996	Fall	10.4	n.a.	-.18
1997	Spring	8.7	n.a.	-.51
1998	Fall	10.1	.17	.18
1999	Spring	8.7	.19	.18

Source: Tabulations of Current Population Survey, Food Security Supplement data.

Table 4.2: Effect of Survey Season on Measured Prevalence of Food Insecurity with Hunger, Regression Analysis Results

				Number of observations =	**5**
				R-squared =	**.963**
				Coefficients and Significance:	
				Season = Fall =	**.639**
				Significance =	**P=.04**
				Time (months) =	**-.0235**
				Significance =	**P=.03**
Case	**Season**	**Time (months)**	**Percentage of Households with Food Insecurity with Hunger**	**Percentage of Households with Food Insecurity with Hunger (Standard Error)**	**Residual**
1995	Spring	0	3.9	n.a.	.05
1996	Fall	17	4.1	n.a.	.04
1997	Spring	24	3.1	n.a.	-.15
1998	Fall	40	3.5	.12	.04
1999	Spring	48	2.8	.07	.10

Source: Tabulations of Current Population Survey, Food Security Supplement data.

5. CALCULATING ITEM CALIBRATIONS AND HOUSEHOLD SCALE SCORES

The food security scale was developed using statistical methods based on the Rasch measurement model. The model assumes an underlying continuum on which both items and households can be located. It assumes that the probability of a household affirming a specific item depends on the relative severity of the food insecurity of the household and that described by the item. For the food security scale, Rasch-based methods are used to determine item calibrations and household scores. An item's calibration represents the point on the scale at which there is a 50 percent probability that any given household at that severity level will affirm the item. Households with higher values on the scale than a particular item's calibration score have more than a 50 percent probability of affirming that item and conversely, those with lower values have less than a 50 percent probability of affirming the item. Item calibrations are calculated based on overall response patterns of all respondents. They are then used to calculate the severity score of households, based on the household's responses to the entire set of items.

There are two Rasch modeling approaches that have been used to calculate item calibrations, marginal maximum likelihood estimation (MML) and joint or unconditional, maximum likelihood estimation (JML). Details on both methods are presented below. Both methods have been used in various phases of developing the food security scale and assessing the consistency of the data with the statistical assumptions underlying the scale. JML methods were used to initially develop the scale and analyze data from the 1995 survey. Similarly, JML was used by the Economic Research Service for all their analyses and for the development of item and household scores published in the *Guide to Measuring Household Food Security, Revised 2000* (Bickel, G. M. Nord, C. Price, et.al, 2000). It served as the basis for the household scores in the 1998 and 1999 public-use data file as well. The MML method was used to reanalyze 1995 data, to analyze 1996 and 1997 data and as the basis for the household scores included in the 1996 and 1997 public-use data files. Here we compare the two methods as applied to the 1998 and 1999 food security supplement data.

Although the two methods are similar, the results produced are not identical. After a general description of Rasch models and a presentation of the results of the 1998 and 1999 estimation process of the food security scale, JML and MML methods are compared in detail and potential reasons for their differing results are explored. The central questions are whether the estimation method used makes any difference in the item calibrations that are obtained and, if so, which method is more appropriate for analyzing food security items and calculating household insecurity scores.

A. Using Rasch Modeling to Measure Food Insecurity

Rasch modeling relies on the assumption that the phenomenon being measured is continuous and can be portrayed as an interval measure. That is, the relative size of the intervals between household severity scores is meaningful, although the zero point is not. It assumes that each household has a score on a latent (unobserved) property that exists on a unidimensional scale. The model further assumes that each item that is used to form the scale is sensitive at a unique level of severity of food insecurity on this same unidimensional continuum. The probability of an affirmative response to any item is a function only of the respondent's level of food insecurity and the item's level of severity. It is assumed that the probability does not depend upon any of the other test items.

The distances between item scores and the ordering of items are meaningful in relative, but not absolute, terms. In other words, Rasch calibrations for a set of items are invariant relative to each other up to a linear transformation. Thus, comparisons of household scores or item calibrations require that the scales both be set to the same zero point. To accomplish this, the scales are adjusted so that the mean of the item calibrations is the same in both scales. The metric used by USDA's food security measurement project is based on a mean item calibration of 7. The size of the interval on the scale also can be made constant, for comparison purposes. The constant (called a scale factor or slope) used in the food security measurement project is 1. For some comparisons, scales may be adjusted so that the standard deviations of items are the same in both scales.

It is important to note that Rasch models do not assign scale scores to respondents with "extreme" response patterns. That is, if a respondent has no affirmative responses, we only know that the respondent's score is below the range measured by these items. Similarly, if the respondent says "yes" to all items, we know that the respondent's score is above the range that can be measured. Thus, in the food security application, households that answered no to all items (raw score of zero) did not have a scale score derived. Neither did households with severe food insecurity who answered yes to all items (raw score of 18). While we understand that the former group is food secure, we do not know how much more secure they are than households that answered only one item positively.

B. Food Security Item Calibrations, 1998 and 1999

Table 5.1 presents the item calibrations calculated for 1998 and 1999 using MML methods implemented by BILOG software[4]. The similarity in relative item severity between the two years as well as to earlier years is evidence of the stability of the measurement construct, and justifies comparison of prevalence rates across years. While in general the magnitudes of the calibrations are ordered similarly, there were two item reversals between the 1998 and the 1999 calibrations. In 1998, the survey question number 8a "adult cut or skipped meals, 3 or more months" had a calibration of 6.482 and question number 7 "children not eating enough" had a calibration of 6.738. By contrast, in 1999, while question number 8a "adult cut or skipped meals, 3 or more months" had an item calibration of 6.497, question number 7 "children not eating enough" had an item calibration of 6.383. This means that in 1998 "children not eating enough" was more severe than "adult cut or skipped meals, 3 or more months", but that the reverse was true in 1999.

There are a number of reasons why this might have happened. Sampling error might be responsible for the change in ordering but given that the differences are on the

[4] A detailed description of procedures used to calculate item calibrations and household scores using BILOG is available from ERS.

TABLE 5.1: 1998 and 1999 Food Security Item Calibration Values [a] (Discrimination (Slope) Parameter Set to 1.0)[b]

Survey Question Number [c]	Item Description	1998 Item Calibration	1998 Standard Error[d]	1999 Item Calibration	1999 Standard Error[d]
2	Worried food would run out	2.14	0.043	2.03	0.046
3	Food bought didn't last	3.40	0.037	3.10	0.041
5	Relied on a few kinds of low-cost foods for children	3.82	0.039	3.67	0.042
4	Couldn't afford to eat balanced meals	4.18	0.047	4.05	0.052
6	Couldn't feed the children a balanced meal	5.36	0.054	5.22	0.061
8	Adult cut size of meals or skipped meals	5.62	0.062	5.71	0.068
9	Respondent ate less than felt they should	5.75	0.049	5.87	0.055
8a	Adult cut or skipped meals, 3 or more months	6.48	0.055	6.50	0.063
7	Children not eating enough	6.74	0.042	6.38	0.048
10	Adult hungry but didn't eat	7.40	0.054	7.45	0.062
11	Respondent lost weight	8.29	0.062	8.41	0.072
13	Cut size of child's meals	8.57	0.088	8.89	0.103
12	Adult did not eat for whole day	8.73	0.113	8.74	0.132
15	Child hungry but couldn't afford more food	8.96	0.096	9.16	0.115
12a	Adult did not eat for whole day, 3 or more months	9.44	0.173	9.46	0.292
14	Child skipped meal	9.56	0.218	9.76	0.328
14a	Child skipped meals, 3 or more months	10.17	0.116	10.04	0.136
16	Child did not eat for whole day	11.33	0.230	11.55	0.281
	MEAN	7.00		7.00	
	STANDARD DEVIATION	2.584		2.695	

[a] Based on August 1998 Current Population Survey Food Security Supplement data.
[b] Adjusted to discrimination parameter (slope) = 1 and mean item score = 7.
[c] Items are ordered by severity reflected by 1998 item calibrations, which is slightly different from the order in the survey and that of 1999 calibrations. Question numbers are those in *Bickel, et al, (2000)* to facilitate comparison across years.
[d] Standard errors do not take account of possible inter-respondent correlation due to cluster sampling.

order of two standard errors, this is not very likely. The second possible explanation is that there is something in either BILOG software, or in the default settings when using that software, that renders item calibrations for BILOG less stable than those for the JML application. (This reversal is not seen in the JML estimates presented later in this section). The differences between the two programs are explored in more detail below. Third, a real change in the perceived meaning of items or their relationship to food security may have occurred.

The second such reversal occurred between “cut size of child’s meals” (question number 13) and “adult did not eat for whole day” (question number 12). In this case, however, the differences are within about one standard error and the reversal may well be the result of sampling variation.

Table 5.2 presents item calibrations based on separate estimations for 1995-1999 calculated using MML methods implemented by BILOG software. Each set of item calibrations has a mean set at 7 with the slope of the item characteristics curves at their inflection points set at 1. Each year’s scores are presented in the severity order reflected by 1998 calibrations. Given that the questionnaire underwent considerable reorganization in 1998, comparisons to that year’s data allow for an assessment of stability not only across years, but also across the questionnaire reorganization.

While the calibrations for all years are relatively similar in magnitude for most items and each scale spans between 8.7 to 9.5 units, there are some changes in the severity of items and differences in the order of severity when each year’s data are compared to 1998 and 1999. While the severity of the first two child items (question 5: “relied on a few kinds of low-cost foods for children” and question 6: “couldn’t feed the children a balanced meal”) remained approximately the same between 1995 and 1997, there was an apparent decline in their severity in 1998 and 1999. This is the result of the greater dispersion of items (higher standard deviation) in 1998 and 1999, reflecting a higher item discrimination. The higher item discrimination resulted from the introduction

TABLE 5.2: Comparison of Item Calibrations Estimated from April 1995, September 1996, April 1997, August 1998 and April 1999 CPS Food Security Data (Discrimination (Slope) Parameter Set to 1.0)[a]

Survey Question Number [b]	Item Description	1995 Item Calibration	1996 Item Calibration	1997 Item Calibration	1998 Item Calibration	1999 Item Calibration
2	Worried food would run out	2.55	2.47	2.37	2.14	2.03
3	Food bought didn't last	3.77	3.74	3.73	3.40	3.10
5	Relied on a few kinds of low-cost foods for children	4.34	4.37	4.53	3.82	3.67
4	Couldn't afford to eat balanced meals	4.02	3.99	4.10	4.18	4.05
6	Couldn't feed the children a balanced meal	5.61	5.62	5.82	5.36	5.22
8	Adult cut size of meals or skipped meals	5.54	5.47	5.56	5.62	5.71
9	Respondent ate less than felt they should	5.63	5.59	5.65	5.75	5.87
8a	Adult cut or skipped meals, 3 or more months	6.35	6.39	6.45	6.48	6.50
7	Children not eating enough	6.95	6.98	6.91	6.74	6.38
10	Adult hungry but didn't eat	7.21	7.20	7.29	7.40	7.45
11	Respondent lost weight	8.28	8.13	8.16	8.29	8.41
13	Cut size of child's meals	8.37	8.57	8.56	8.57	8.89
12	Adult did not eat for whole day	8.47	8.46	8.51	8.73	8.74
15	Child hungry but couldn't afford more food	8.63	8.98	8.95	8.96	9.16
12a	Adult did not eat for whole day, 3 or more months	9.02	9.05	9.01	9.44	9.46
14	Child skipped meal	9.65	9.67	9.29	9.56	9.76
14a	Child skipped meals, 3 or more months	10.19	10.15	9.86	10.17	10.04
16	Child did not eat for whole day	11.44	11.15	11.27	11.33	11.55
	MEAN	7.00	7.00	7.00	7.00	7.00
	STANDARD DEVIATION	2.449	2.450	2.40	2.584	2.695

[a] Adjusted to discrimination parameter (slope) = 1 and mean item score = 7.

[b] Items are ordered by severity reflected by 1998 item calibrations, which is slightly different from the order in the survey and that of 1995,1996,1997 and 1999 calibrations. Question numbers are those in *Bickel, et.al, (2000)* to facilitate comparison across years.

of internal screens and reordering of items in 1998, both of which tend to increase the consistency of response with item severity[5].

As the estimated severity level of these two items decrease, there is a resulting inversion in the severity ranking of the items following them. That is, in 1995, 1996 and 1997, question 5 ("relied on a few kinds of low-cost foods for children") was ranked as being more severe than question 4 ("couldn't afford to eat a balanced meal"). In 1998 and 1999 this ordering is reversed. In 1995-1997 the item calibrations for "couldn't afford to eat balanced meals" were lower than "relied on a few kinds of low-cost foods for children" by 0.326, 0.389 and 0.427 units respectively. In 1998 and 1999 the order of severity is reversed with item calibrations for "relied on a few kinds of low-cost foods for children" being lower than "couldn't afford to eat balanced meals" by 0.360 and 0.372 units respectively.

A similar pattern is observed between questions 6 and 8. In 1995-1997, item calibrations for question 6 "couldn't feed the children a balanced meal" were higher than those for question 8, "adult cut the size of meals or skipped meals". The opposite occurred in 1998 and 1999. Unlike the reversal described above however, the magnitude of the difference increased consistently through the years from .072 in 1995 to .494 in 1999. In 1996 and 1997 question 6 also fell below question 9 ("respondent ate less than they felt they should") in severity order.

It is most likely that these changes in relative item severity are the result of changes in the order of administration of the questions in 1998 and 1999. An exploration of the effects of screening on item calibrations (not shown here) has revealed that the internal screens increase the dispersion (standard deviation) of items by approximately three percent, but the effects of screening (initial, common, or internal) on relative severity of items are negligible and do not contribute in any substantial way to the

[5] Technically, the internal screens introduce inter-item dependencies that violate Rasch model assumptions. However, analysis of data from 1995-1997, which was not affected by internal screening, indicates that screening has negligible effects on relative item severities.

reversals of interest. Overall, the similarity of most item calibrations across the years and the consistency in their order of severity is indicative of the stability of the measurement construct across the years and across the significant reorganization of the questionnaire in 1998.

C. Comparing MML to JML Procedures

Item calibrations produced by BILOG's MML procedures are very similar to those produced by JML procedures[6] but are not identical. They differ somewhat because the conditions that the two methods impose on the maximum likelihood solution differ. Further, characteristics of the BILOG software in conjunction with characteristics of the food security data limit the precision of the MML estimates and require special handling. These differences do not threaten the meaning or reliability of the measure. The following compares the item calibrations developed by the two procedures and explores factors contributing to these differences.

To facilitate comparisons of item scores obtained using BILOG (MML) and JML procedures, item discrimination (or item slopes) must first be adjusted so that the dispersion of item scores from the two procedures are equal. With item slopes set to 1.0, as in Tables 5.1 and 5.2, the dispersion (as measured by standard deviation) of the BILOG item calibrations for the 18 items is less than that of JML calibrations. For 1998, the BILOG standard deviation is 2.584 and the JML standard deviation is 2.996, while for 1999 the standard deviations are 2.695 and 3.145 respectively.

The larger standard deviation of the JML estimates is consistent with a known upward bias on the dispersion of JML item calibrations. JML item calibration estimates are not statistically consistent. That is, as the sample size increases without limit, the JML item calibrations do not converge to their expected values. They are biased toward greater dispersion than the "true" item calibrations. However, the direction and approximate size of the bias is known. It has been shown that the JML inconsistency bias

---

[6] Joint maximum likelihood (JML) calibrations were calculated using an ERS adaptation of WINSTEPS (Lineacre and Wright, 1998). The ERS adaptation allows use of household case weights for estimating item parameters.

requires multiplication by a correction factor, (L-1)/L, to approximate consistency where L is the number of items (Andrich, 1988). Notice, however, that this would account for only about half of the difference observed in this case. Andrich's correction would be 17/18 (about -6 percent), or for households without children it would be 9/10 (-9 percent), whereas the observed differences in standard deviations are about 16 percent.

Because of the difference in dispersions of item parameters as estimated by BILOG and the JML method, we adjust to an equal standard deviation prior to an item-by-item comparison of severity[7]. Table 5.4 presents a comparison of the item calibrations calculated for 1998 and 1999 by JML procedures with the adjusted BILOG calibrations. While some of the differences between the two estimates are small, some are modestly substantial. For example, in the 1998 calibrations, four items have a difference between 0.10 and 0.16 including: "adult did not eat for a whole day for 3 or more months," "worried food would run out," "respondent lost weight," and "adult did not east for a whole day." In 1999 these four items continue to have a modestly substantial difference between the calibrations (0.13-0.17), along with two additional items which have a difference of –0.17 ("adult hungry but did not eat" and "child did not eat for whole day"). For both years the differences are greatest for the extreme items, either those that are most or least severe.

With a single exception, the ordering of the items is invariant between the two methods within each year. For the 1999 estimates, the JML procedures give the calibration of "adult did not eat for whole day" as 9.19 and the calibration of "cut size of child's meals" as 9.18, while BILOG gives the adjusted item calibration for "adult did not eat for whole day" as 9.02 and the calibration of "cut size of child's meals" as 9.19.

[7]BILOG scores were adjusted by a linear transformation to result in the same mean and standard deviations as the JML scores.

TABLE 5.4 Difference Between BILOG and JML Item Calibrations [a]

Item Description [b]	Item Calibration		JML – BILOG Difference 1998	Item Calibration		JML - BILOG Difference 1999
	JML 1998	BILOG 1998		JML 1999	BILOG 1999	
Worried food would run out	1.49	1.37	0.11	1.31	1.18	0.13
Food bought didn't last	2.79	2.83	-0.04	2.41	2.43	-0.02
Relied on a few kinds of low-cost foods for children	3.27	3.32	-0.05	3.03	3.10	-0.07
Couldn't afford to eat balanced meals	3.67	3.74	-0.07	3.48	3.54	-0.06
Couldn't feed the children a balanced meal	5.04	5.10	-0.06	4.83	4.90	-0.07
Adult cut size of meals or skipped meals	5.37	5.41	-0.03	6.22	6.26	-0.04
Respondent ate less than felt they should	5.53	5.56	-0.03	5.68	5.67	0.01
Adult cut or skipped meals, 3 or more months	6.42	6.41	0.02	6.46	6.40	0.07
Children not eating enough	6.66	6.70	-0.04	6.22	6.26	-0.04
Adult hungry but didn't eat	7.55	7.47	0.08	7.65	7.83	-0.17
Respondent lost weight	8.61	8.51	0.11	8.79	8.62	0.17
Cut size of child's meals	8.79	8.82	-0.03	9.18	9.19	-0.01
Adult did not eat for whole day	9.12	8.96	0.16	9.19	9.02	0.17
Child hungry but couldn't afford more food	9.24	9.27	-0.03	9.49	9.50	-0.01
Adult did not eat for whole day, 3 or more months	9.93	9.83	0.10	10.01	9.86	0.15
Child skipped meal	9.94	9.97	-0.04	10.17	10.20	-0.03
Child skipped meals, 3 or more months	10.63	10.68	-0.05	10.49	10.53	-0.04
Child did not eat for whole day	11.94	12.03	-0.09	12.12	12.29	-0.17
MEAN	7.00	7.00		7.00	7.00	
Standard Deviation (SD)	2.996	2.996		3.145	3.145	
Original SD ratio	1.16				1.17	

[a] Based on August 1998 and April 1999 Current Population Survey Food Security Supplement data. JML metric is in logistic units. BILOG metric is adjusted so that the standard deviation of the BILOG item scores is the same as that of the JML item scores.

[b] Items are ordered by severity reflected by 1998 item calibrations, which is slightly different from the order in the survey and that of 1999 calibrations.

Differences in the ways in which BILOG and JML applications employ maximum likelihood techniques cause these small differences in calibration results. The following explores some specific differences in program characteristics that may explain why the results vary.

Conceptual Differences

Conceptual differences between MML and JML procedures provide some explanation as to the variation in results. MML methods estimate the probability that a person with a particular score obtains a positive answer to a particular item with the individual person parameters conditioned away (or integrated out) (Andrich, 1988). This is based on an assumption about the distribution of the severity in the population. It is only dependent upon the values of all the item parameters for estimation and the assumptions about the distribution in the population. This means that the MML solution assumes that respondents are drawn randomly from a population of severities that is either a normal distribution or an arbitrary smooth distribution specified by the user (BILOG, 1990).

The JML procedure estimates both household and item parameters simultaneously. No assumption is required about distribution of the trait in the population. Household and item scores are estimated so as to maximize the likelihood of getting the observed matrix of responses under Rasch assumptions. The process begins with rough estimates of scores for each respondent and each item. These are then refined through an interactive Newton-Raphson process toward maximum likelihood until further parameter adjustments are smaller than specified convergence criteria.

Andrich comments that while the unconditional JML procedure is more efficient and converges faster than the MML procedure used by BILOG, it produces inconsistent estimates (in a statistical sense) for all of those estimates that are based upon a fixed number of items (questions). Statistical inconsistency is, ideally at least, undesirable in a measure because this means that even if the item and sample size is increased without limit, the estimate remains biased and does not converge to the value of the population parameter. This bias increases the dispersion of item scores estimated by JML methods relative to the dispersion of item scores estimated by MML methods. This is exactly what was found in the JML - BILOG comparisons described above. The size of this bias is known approximately, and is not generally problematic for the food security scale. It should, however, be kept in mind when comparing results from scales based on a subset of the items, such as the six-item standard short module.

Problems Introduced by MML Assumption of Smooth Distribution

As estimated under MML assumptions, the severities of the child item scores (or some of them, at least) are distorted relative to those of the adult and household item scores because of the assumption inherent in the MML method of smooth distribution of severity in the population. The distortion arises because in fact, two populations—households with and without children—are intermingled in the food security data. This wouldn't matter if all households got the same set of items, but because all of the child items are estimated based only on the population of households with children, the two sets of household score groups are interspersed, but represent the two populations. (And, the more severe household score groups are all based only on households with children.) Depending on the exact method used to assess the distribution, BILOG may adjust child items disproportionately relative to other items to try to smooth the distribution. BILOG (MML) and JML procedures produce item calibrations that are almost perfectly linear with respect to each other if the universe is restricted to households with children or to households without children. However, when the population is mixed (households with and without children), the calibrations are no longer perfectly linear. Thus, the BILOG (or generic MML) assumptions of a smooth distribution (inappropriate when the two household types are mixed) account for most of the non-linearity between BILOG and JML item scores described above. The MML assumption regarding distribution of the measured trait in the population may also be violated by the screening procedures in the Food Security Supplement. This violation may also contribute to the difference between JML and MML scores. Even so, the differences are not very large and in most cases have little practical implication. But they are large enough to be noticeable.

Convergence

The major problem with using BILOG to estimate item calibrations from the food security data is lack of convergence in the Newton Raphson phase. The non-convergence problem is a result of the dependent frequency follow-up items. Those items and their base items oscillate in opposite directions in alternate steps of the iteration process. The size of the oscillation is quite large, in the range of 2 to 4 logistic units. If the frequency follow-ups are removed, then the calibrations will converge to 0.001 in 5 to 15 NR steps. Similarly, if the base items of the frequency follow-ups are removed and their frequency follow-ups included, the calibrations will converge.

In an attempt to resolve this problem, the free option (an option on the Calibration command) was tested. It was thought that use of an empirically derived prior distribution as opposed to a program-imposed normal distribution might resolve the convergence problem. Although the results are not shown here, these calibrations also failed to converge on a stable estimate. To address the issue of non-convergence, the two item sets were estimated separately (with frequency follow-ups excluded in one run and their base items excluded in the second), and then the metrics were equated by the common items. Interestingly, the set of scores derived by this process is very nearly linear with the scores at the end of the so-called EM estimation phase (after 10 cycles or so). Thus, the item dependencies appear to be problematic only in the Newton-Raphson phase, not in the EM phase.

It is also worth pointing out that the BILOG software and manuals are oriented to item response theory (IRT) users in general and not to Rasch modelers in particular. This makes it more difficult for the novice user to use BILOG to estimate a one-parameter logistic (Rasch) model than a JML program such as WINSTEPS, which is specifically oriented to Rasch's own views of measurement rather than IRT more generally.

D. Households Scores, 1998 and 1999

Household scores on the food security scale are also calculated based on the Rasch model. One result of the Rasch assumption that items discriminate equally is that, for a given set of items, the household's scale score depends only on the number of items affirmed, not on which items are affirmed. Thus, for households with no missing data, households with the same raw score will be assigned the same scale score. If respondents are not all given the same set of questions, the scale scores depend on the severity (as indicated by the item calibration) of the questions that the respondent answers, as well as the number of items affirmed. The food security supplement includes 18 questions for household with children and 10 questions for those without. The Rasch model takes these differences into account, assigning scores to both types of households that are comparable even though they responded to different subsets of questions. The model also adjusts the household scale scores for households that failed to respond to one of more of the applicable questions.

Once item calibrations have been determined, household scale scores for households with no missing values can be calculated for each raw score. Identical household scores are produced by BILOG and JML software provided the same item calibrations are specified. After reviewing all item calibrations, it was determined that the JML calibrations would be used as the "standard". Although JML estimates are somewhat biased toward greater dispersion than their true value, ordinality of items is preserved and relative severities (proportional size of intervals between items) appear to be either unaffected or minimally affected by the statistical inconsistency. These latter two important characteristics cannot be assured in MML estimates except through a complex process of fitting multiple models. Table 5.5 presents the household scores that were used for both 1998 and 1999 public-use data files, which were based on JML methods and the 1998 data.

Table 5.5 Household Food Security Scale Scores, 1998 and 1999

Number of "yes" responses		Household Scale Scores 1998 and 1999	Food Security Status Category
Household with child	Household without child		
1		1.428	Food Secure
	1	1.723	
2		2.560	
	2	3.101	
3		3.405	Food Insecure without Hunger
4		4.138	
	3	4.232	
5		4.138	
	4	5.234	
6		5.430	
7		6.024	
	5	6.155	
8		6.606	Food Insecure with Hunger
	6	7.068	
9		7.179	
10		7.738	
	7	8.002	
11		8.276	
12		8.794	
	8	8.976	
13		9.306	
14		9.837	
	9	10.149	
15		10.423	
16		11.133	
17		12.157	

REFERENCES

Anderson S., (ed.). "Core indicators of nutritional state for difficult-to-sample populations." A report prepared by the Life Sciences Research Office, Federation of American Societies for Experimental Biology for the American Institute of Nutrition under Cooperative Agreement No. HPU 880004-01-0, Nutritional Status Indicators of Low-Income Populations, with the Office of Disease Prevention and Health Promotion, Department of Health and Human Services. Journal of Nutrition, 120 (11S): 1557-1600, 1990.

Andrews, M., Nord, M., Bickel, G., and Carlson, S. "Household Food Security in the United States, 1999 Food Assistance and Nutrition Research Report, no. 8." Washington, DC: Economic Research Service, U.S. Department of Agriculture, 2000.

Andrich, D. "Rasch Models for Measurement." Newbury Park, CA: Sage Publications, 1988.

Bickel, G., Carlson, S., and Nord, M. "Household Food Security in the United States, 1995-1998, Advance Report." Alexandria, VA: Office of Analysis, Nutrition and Evaluation, Food and Nutrition Service, U.S. Department of Agriculture, 1999.

BILOG 3, 2nd Edition. Chicago, IL: Scientific Software, Inc. 1990.

Campbell C. C. "Food insecurity: a nutritional outcome or a predictor variable?" Journal of Nutrition. Mar; 121(3): 408-15. 1991.

Cohen, B., Lerner, R., and Parry, J. "Household Food Security in the United States, 1998 and 1999: Technical Report." Food Assistance and Nutrition Research Report. Washington, DC: Economic Research Service, U.S. Department of Agriculture, 2002.

Cohen B., and Burt M. "Eliminating hunger: food security policy for the 1990s." Washington, D.C.: The Urban Institute, 1989.

Hamilton W., Cook J., Thompson W., Buron L., Frongillo E., Olson C., Wehler C. "Household food security in the United States in 1995, Summary Report of the Food Security Measurement Project." Alexandria, VA: Food and Consumer Service, U.S. Department of Agriculture, July 1997.

Hamilton W., Cook J., Thompson W., Buron L., Frongillo E., Olson C., Wehler C. "Household food security in the United States in 1995, Technical Report." Alexandria, VA: Food and Consumer Service, U.S. Department of Agriculture, July 1997.

Lineacre, J., and B. Wright. "A user's Guide to BIGSTEPS/WINSTEPS Rasch Model Computer Program." Chicago: MESA Press (University of Chicago). 1998.

Ohls, J., Radbill, L., and Schirm, L. "Household Food Security in the United States, 1995-1997: Technical Issues and Statistical Report" Alexandria, VA: Food and Consumer Service, U.S. Department of Agriculture, December, 2001.