PSID Technical Report
The 2002 PSID Child Development Supplement (CDS-II) Weights

Elena Gouskova, Ph.D.
August 24, 2001

1.   Introduction

Child Development Supplement (CDS) to the Panel Study of Income Dynamics (PSID) is a longitudinal study that began in 1997. It was designed to provide researchers with comprehensive nationally representative data about children ages 0-12 and their families. All respondents to the Child Supplement have been selected from the PSID families.

In the first wave of the CDS 2,394 families participated, providing information on a total sample of 3,563 children. The second round of data collection took place in 2002-2003 when 2006 families were successfully re-interviewed, resulting in total of 2907 child interviews. The reduction in the sample is due to reclassification of the eligibility status of a number of 1997 sample participants and also due to nonresponse among the remaining eligible CDS families.

   To account for differential probabilities of selection due to the original PSID sample design and subsequent attrition, CDS data are provided with weights. This note describes the process of the 2001 Child Supplement weight construction and discusses some related issues.

The note is organized as follows. Part two provides information on the sample attrition between 1997 and 2001. The Primary Caregiver/Child weight construction procedure is described in Part 3. In Part 4 we discuss the complex structure of the CDS which consists of several interview modules. In this section we highlight the issue of missing data for special modules and suggest possible ways to deal with this missing data problem in the analysis of these data.

2.   Changes in the CDS Sample Between 1997 and 2002/2003

Out of the 3,563 children with complete CDS-I interviews in 1997, information was gathered for 2907 children in CDS-II during 2002/2003. The reduction in the sample comes from two sources: the decision not to follow some families from the original 1997 CDS sample (292 children) and non-participation by eligible families in the 2002-2003 interview process (364 children). A summary of the sample size changes in the CDS panel is provided in Table 1.

Prior to the 2002-2003 CDS data collection, a decision was made to follow only 1997 CDS baseline families with children that had sample status[1]. Of the 3,563 children who participated in CDS-I, 83 were "non-sample" according to standard PSID rules for longitudinal panel eligibility for newborns and other new household members.

PSID respondents are distinguished as sample and non-sample members. Sample respondents are those who belong to the original 1968 family for the core PSID sample or original 1997 family for Immigrant sample. Sample members are also those who related by blood or adoption to a PSID sample member.[1]

The second eligibility criteria for CDS-II was that the family of the targeted CDS child needed to be active in the PSID panel as of the 2001 main interview. Between 1997 and 2001, 94% of the CDS children had families who remained active in PSID (that is participated in the 2001 PSID biennial interview). Thus, the suspended portion of the sample consisted of families with non-sample children and families that left the PSID study before 2001. The total number of suspended cases is 292, reducing the targeted sample of children from 3563 in 1997 to 3271 who were eligible to participate in 2002/2003.

Starting with this sample, the CDS-II successfully obtained interviews with 2006 PSID families, resulting in 2907 child observations. Accounting for both sources of attrition in the 1997 baseline CDS-I panel, the CDS-II re-interview rate is 81.6%. Among PSID "sample" children determined to be eligible for CDS-II re-contact, the interview completion rate for the primary care giver (PCG) interview was 88.9%.

3.   CDS-II Weight for Primary Caregiver (PCG)/Child Analysis

CDS-II, like CDS-I before it, includes different interview modules that are designed to collect data from multiple "observers" that may be associated with each sample child. In CDS-II, these "observers" always include the child's primary caregiver (PCG) but may also include: 1) a secondary care giver; 2) a teacher; and 3) the child herself/himself (both interview and test assessment). The primary caregiver interview is the minimum data required to qualify as a CDS-II case. Table 1 summarizes the eligibility and response outcome for the PCG interview. Eligibility and response dispositions for the remaining CDS observation modules are summarized in Table 2.

The CDS-II analysis weight described below pertains to the analysis of the PCG interview data. Special weights are not provided for the analysis of the other CDS-II modules listed above. Section 4 below provides guidance on how analysts may address the problem of missing data for other child observation modules.

Sample survey data are typically provided with weights designed to compensate for unequal probabilities of sample selection and non-response or data that is missing at random (MAR) (Little and Rubin, 2002)[2]. These weights are inversely proportional to the probability that each observation is selected and conditional on selection that they respond to the survey questions. With longitudinal data this joint probability at time t, where the study has started at t-1 or earlier, can be expressed as following

(1) $P(S_t=1)=P(S_{t-1}=1)*P(R_t=1|S_{t-1}=1)$

where S_t is an indicator of participation in a study at time t and R_t is an indicator of response at time t. The probability of being a participant at time t is a product of probability of participating in previous period and conditional probability of being a response in the current period. As the first term on the right hand side of (1) is proportional to a reciprocal of weight in the previous period, the weight in the current period is a product of the weight in previous period and the inverse of probability of response (the second term on the right hand side of (1)). We will refer to $1/P(R_t=1|S_{t-1}=1)$ as attrition adjustment factor.

Accordingly, the CDS-II weight is a product of CDS-I weight and the attrition adjustment factor. The individual level Primary Caregiver/ Child weight (CH97PRWT) was used as the previous year weight[3].

---

[2] Little, R.J.A. and Rubin, D.B. (2002). Statistical Analysis With Missing Data, 2nd Edition. John Wiley & Sons, New York.

[3] For the description of the 1997 CDS weights construction see
http://psidonline.isr.umich.edu/CDS/weightsdoc.html

To obtain the attrition adjustment factor the probability that a sample person was successfully re-interviewed in the CDS-II was modeled with the linear logistic model. As we discussed above there were two different reasons for attrition. We modeled them together, i.e. observations where a person was not interviewed included both suspended and non-response cases[4]. The estimates of the logistic model are given in Table 3.

As the final step in the weight development, the newly constructed CDS-II weights for Primary Caregiver/child observations were censored to reduce the influence of extreme weights on the variances of sample estimates of population statistics. One percent of weights at the top and bottom of the distribution were assigned values of the 99th and 1st percentiles respectively.

To check the 2002/2003 CDS-II PCG weight we compared weighted estimates for some basic demographic, geographic and socio-economic variables in 1997 (weighted by the 1997 CDS-I weight) to the same estimates for the smaller CDS-II re-interview sample (weighted by the CDS-II weight). The results of this comparison are provided in Table 4. The comparison of weighted estimates for the baseline CDS-I and follow-up CDS-II suggest that the CDS-II attrition adjustment factors included in the PCG/Child analysis weights are compensating for potential attrition bias in the family type and demographic composition of the CDS panel data. Note: This comparison does necessarily not rule out the possibility of spurious or more subtle forms of selection bias that may not be associated with the demographic, geographic and socio-economic characteristics of CDS respondents.

The final CDS-II weight is stored in the variable CH02PRWT. The CH02PRWT weight should be used for analyses involving child level data or data involving the relationship of the child with their PCG or with family characteristics.

The Primary caregiver/ Child weight, CH02PRWT, can be converted into a caregiver level weight by averaging the weights for each CDS-II child observation provided by a caregiver. This PCG/family weight could be use to develop an analysis that is focused on the caregiver or their family as the unit of observation/analysis.

4.   Addressing Missing Data in Other CDS-II interview Modules

The CDS interview has a complex structure involving several modules. These modules are the Primary Caregiver (sections A-H, J), Child Interview, Child Assessment, Secondary Caregiver, and Elementary School Teacher Interview. The Primary interviews form the core data collection of the Supplement. Completion of these interviews determined whether an observation is included in the final data set or not. Thus for all children in the CDS data set the Primary Caregiver interview were responses. Response /non-response the Primary caregiver interview was also an indicator for attrition in calculation of the CDS-II attrition adjustment parameter described in the previous section.

Other CDS interview modules are secondary to the Primary Caregiver module and may or may not have been successfully completed. It is possible to have information on the Primary Caregiver interview and have missing data for any of the secondary module interviews. It is important to note that all secondary interviews, except for the Child Assessment interview, had an additional eligibility condition. For example, the Child Interview was administered only to children ages 8 or older. Thus missing data on the secondary interview with eligibility condition can be due to two reasons: non-response or ineligibility.

---

[4] We experimented with separate modeling of the two types of attrition. The results turned out to be very similar.

When accounting for the eligibility restrictions, the response rate to the secondary interviews is generally lower than the response to the Primary Caregiver interview. The case counts by eligibility and response status for each secondary module are given in Table 2.

Researchers interested in analyzing the CDS-II data for the secondary modules should carefully examine the extent of the partially missing data problem and address it if necessary.

The issue can be approached in a number of ways. The simplest way is to assume that the data are missing completely at random (MCAR). That is, given the adjustment that has already been performed for attrition in the CDS PCG observation sample, non-response to a secondary module is completely random with respect to the dependent variables of interest. When MCAR is the appropriate assumption, analysis of the observed data for the secondary modules will not lead to biased estimation. The MCAR assumption, however, is too strong and may not always be a good approximation for an attrition process. For example, our preliminary investigation indicates that for the Child Interview, age was an important factor in participation decision. The younger children tend to be more likely to participate than older children. If the probability of nonresponse can be modeled as a function of observed variables such as age or gender, Little and Rubin (2002) describe the missing data mechanism as missing at random (MAR).

When the nonresponse or missing data mechanism is MAR, one can approach the problem by either imputing the missing values or constructing additional weighting adjustments using procedures that take into account variables (such as age in the above example) that predict the probability of a response. Analysts interested imputation of missing response data for the CDS secondary modules may wish to review Raghunathan, et al. (2001)[5] and access the corresponding SAS-compatible imputation software at http://www.isr.umich.edu/src/smp/ive/. However, since the missing data problem for CDS-II secondary modules typically involves complete missing data for all variables, analysts may choose to develop a weighting adjustment to compensate for an MAR missing data mechanism. The algorithm for constructing these weighting adjustments should be similar to the one described above in Section 3 for the PCG/Child. Here, the logistic regression model of response propensity would be estimated based only on characteristics of eligible respondents and nonrespondents, excluding cases that were not eligible for the given module (see Table 2). The new attrition adjusted weight would be constructed as the product of the individual weight (CH02PRWT) and reciprocal of the estimated probability of response in the secondary interview.

---

[5] Raghunathan, T.E., Lepkowski, J.M., VanHoewyk, J., and Solenberger, P. (2001). "A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models". Survey Methodology: Vol. 27, No.1. pp. 85-95. Statistics Canada.

Table 1

Suspended in 2001		292
	Subtotal	***292***
Targeted in 2002	No PCG interview	364
	Yes PCG interview	2907
	Subtotal	***3271***
Total (1997 sample size)		***3563***

Table 2

Questionnaire Module	Status	N	Response Rate*
Child Interview	Not eligible	251	
	Yes interview	2176	
	No interview	480	
	Total	2907	81.9%
Child Assessments	Not eligible	0	
	Yes interview	2644	
	No interview	263	
	Total	2907	91.0%
Secondary Caregiver	Not eligible	898	
	Yes interview	1686	
	No interview	323	
	Total	2907	83.9%
Elementary School Teacher Interview	Not eligible	1602	
	Yes interview	699	
	No interview	606	
	Total	2907	53.6%

*Excluding ineligible cases.

Table 3.

Variable	Estimate	Std Err	Wald ChiSq	Prob ChiSq
Intercept	-0.2847	0.3664	0.60	0.4372
D Sample	-0.9745 ***	0.1204	65.51	<.0001
D SRC sample	-0.6414 ***	0.228	7.91	0.0049
D SEO sample	-0.5121 **	0.246	4.33	0.0374
D Head and Wife in FU	-0.9601 ***	0.2131	20.29	<.0001
D Child is male	0.0804	0.0889	0.82	0.3655
Age of child at 1997 PSID core interview	0.0336 **	0.0133	6.39	0.0115
D White (child race)	-0.1442	0.2028	0.51	0.477
D Black (child race)	-0.447 **	0.2134	4.39	0.0362
D age<=30 (head)	0.1447	0.1603	0.81	0.3667
D 30< age<=45 (head)	-0.0411	0.1374	0.09	0.7648
D Head is male	0.6654 ***	0.2226	8.94	0.0028
D No High School Degree (head)	0.2276	0.1596	2.03	0.1538
D High School Degree (head)	0.0872	0.1452	0.36	0.5482
D Some College (head)	0.1417	0.1499	0.89	0.3446
D Head is employed	-0.2221 *	0.1279	3.02	0.0824
D Inc in 1st quartile	-0.3431 *	0.1807	3.61	0.0576
D Inc in 2nd quartile	-0.1062	0.1506	0.50	0.4807
D Inc in 3rd quartile	-0.088	0.137	0.41	0.5206
D Northeast	0.5613 ***	0.1723	10.61	0.0011
D North Central	0.4618 ***	0.1581	8.53	0.0035
D South	0.603 ***	0.1493	16.32	<.0001
D MSA 1990	0.0365	0.1104	0.11	0.7407
Summary statistics				
N:	3563			

Table 4.

		Weight 1997		Weight 2001	
		N	**perc**	**N**	**perc**
Total		3563	100.0	2907	100.0
Region	Northeast	483	18.0	389	17.8
	North_Central	877	24.1	720	23.8
	South	1599	33.4	1283	33.0
	West	604	24.5	515	25.4
Immigrant sample	Non-IMM	3234	83.9	2659	83.8
	IMM	329	16.1	248	16.2
MSA	Non-MSA	867	29.6	710	30.0
	MSA	2696	70.4	2197	70.0
Education of Head	No High School Diploma	806	19.0	630	18.8
	High School Diploma Only	1284	32.2	1057	32.1
	Some College or More	797	23.3	650	22.9
	College or More	676	25.4	570	26.1
Age of Head	30 or younger	912	22.0	735	22.2
	31-45	2208	66.2	1820	66.1
	46 or older	443	11.8	352	11.7
Gender of Head	Female	1063	22.1	852	21.4
	Male	2500	77.9	2055	78.6
Race of Head	Non-Black	2102	84.1	1721	84.8
	Black	1461	15.9	1186	15.2