

Response to the Office of Management and Budget

2011 California Health Interview Survey (CHIS 2011)

OMB #: 0925-0598

Would you please ask the Program if there is an update regarding section B2.2. copied below?

B.2.2. Estimation Procedure

CHIS data will be statistically weighted to account for the differential probability of selecting persons into the sample, and the weights will be raked to the various domains of California population totals. Specific strategy for drawing and weighting the CHIS sample will depend on the results of the recently fielded CHIS 2009, and as such, final specifications are not available at this time. The methods detailed below, therefore, enumerate the anticipated estimation procedure.

As in the two preceding CHIS data collection cycles (CHIS 2007 and 2009), CHIS 2011 will again deploy multiple sample frames, including a landline RDD sample, a cellular RDD sample, and surname list samples to enhance the participation of Korean and Vietnamese households. The weighting procedure for CHIS 2011 will conceptually follow that of previous CHIS cycles: a standard design-based multiple-frame methodology that is consistent with the sampling methods used. The multiple-frame approach will be used to combine and weight the landline, surname list samples, and cell phone sample and the landline and surname samples.

Following the creation of the household base weight, further adjustments will be made for unknown residential status, list sample eligibility, screener-interview nonresponse, and multiple telephone numbers. The final step is to create weights for the extended adult, child, and adolescent samples from the adjusted household base weights; in this final step, additional adjustments are made for probability of selection, nonresponse, telephone type adjustment, composite weight adjustment for combining the landline and cellular phone samples, and raking adjustments to person-level control totals.

The CHIS 2011 cellular phone sample will follow the CHIS 2009 method. The CHIS 2007 and 2009 cellular phone sample differed in two important ways. First, in the CHIS 2007 cellular sample, a screener interview was employed to limit extended interviews to those without a landline telephone (i.e., to produce a “cell-phone only” sample). In

contrast, CHIS 2009 included all eligible cases from the cellular phone sample, resulting in overlapping cellular and landline RDD frames. Second, the CHIS 2009 cellular sample implementation included child and teen extended interviews whereas children and teens were ineligible in the CHIS 2007 cellular sample. The CHIS 2007 weighting methodology is well documented and publicly available at:

http://www.chis.ucla.edu/pdf/CHIS2007_method5.pdf

Below, we describe the CHIS 2009 approach to weighting the combined and overlapping landline and cellular RDD samples that will be used in CHIS 2011.

Unlike CHIS 2007 where the cell sample was screened and data collected from cell-only households, CHIS 2009 collects information from all households identified in the cell sample. These include cell only adults, adults that receive most of their calls in their cell phone, adults who receive some of their calls and some on their landlines, and adults who received few or no calls on their cell phone. The same groups are also identified in the landline sample if the adult has a cell phone.

Since the landline and cell phone populations and samples overlap and the drawn samples are probability samples, we can use a multiple frame estimation approach to combine and create weights for these samples. This approach follows the ideas of Hartley (1962) and is different from the approach used to combine the landline and surname samples. The proposed method is needed because we cannot determine the multiple probabilities of selection of all units in the sample. The method we plan to use is outlined in Brick, Flores Cervantes, Norman, and Lee (submitted).

There are three domains of interest in the overlapping frames. The first domain called *a* includes all adults in households with only landline service, the second domain called *b* includes all adults in cell-only households, and the third domain called *ab* includes all adults in households with both landline and cell phones. Let Y be a characteristic for adults in a domain (e.g., the number of adults with health insurance). Let \hat{Y}^A be the estimate of Y computed using the landline sample, and let \hat{Y}^B be the estimate of Y computed using the records in cell phone sample. An estimate of Y using the landline sample is

$$\hat{Y}^A = \hat{Y}_a^A + \hat{Y}_{ab}^A,$$

where \hat{Y}_a^A is the estimate computed using the records from landline only households and \hat{Y}_{ab}^A is the estimate computed using the adults with a landline and cell phone from in the landline sample. In a similar way, an estimate of Y based on the cell phone sample is $\hat{Y}^B = \hat{Y}_{ab}^B + \hat{Y}_b^B$ where \hat{Y}_{ab}^B is the estimate computed using the adults with a landline and cell phone from the cell phone sample and \hat{Y}_b^B is the estimate computed using the records from cell only households.

Notice that neither \hat{Y}^A nor \hat{Y}^B are unbiased estimates of Y . However, an unbiased estimate of Y can be computed as

$$\hat{Y} = \hat{Y}_a^A + \lambda \hat{Y}_{ab}^A + (1 - \lambda) \hat{Y}_{ab}^B + \hat{Y}_b^B,$$

where λ ($0 \leq \lambda \leq 1$) is the composite or weighting factor. The value of λ was chosen to minimize the bias of \hat{Y} .

Before creating the composite weights, both samples (landline and cell) were poststratified separately to control totals defined by telephone service or type (i.e., persons in landline only households, persons in cell phone only households, and persons in households with both services). The distribution of telephone usage for California was derived from the National Health Interview Survey for January to June 2010 for the West region. The poststratified person weight, $^{PPERW}_j$ is computed as

$$^{PPERW}_j = \frac{TEL_USAGE_CT_i * PERW_j}{\sum PERW_j}$$

where $PERW_j$ is the person weight (i.e., adult, child, or adolescent) and $TEL_SERVICE_{ct_i}$ is the control total by telephone service.

Once the samples were poststratified, a composite weight that combined the landline and cell phone sample was created. Based on the research by Brick et. al (submitted), using the composite factor $\lambda=0.9$ was used to reduce the bias of estimates computed using both samples. This factor and its complement ($1 - \lambda$) can be seen as an additional weighting adjustment factors to apply to the poststratified weights. The expression of the composite weight, $^{COMBW}_j$, is

$$COMBW_j = \begin{cases} PPERW_j & \text{If person } i \text{ lives in a household with cell only or landline only} \\ \lambda * PPERW_j & \text{If person } i \text{ lives in a household with cell and landline from the landline sample} \\ (1 - \lambda) * PPERW_j & \text{If person } i \text{ lives in a household with cell and landline from the cell sample} \end{cases}$$

where $PPERW_j$ is the poststratified person weight above. The total sum of weights is an estimate of the total eligible population in California.

Since the landline and cell phone samples are independent samples, the estimates of variance can be computed using replication or linearization (i.e., Taylor series approximation).

In summary, the supplemental samples (i.e., geographic and surnames samples) are combined with the landline sample at the beginning of the weighting process. The cell phone sample and the combined landline-supplemental samples are first poststratified to telephone status or usage, combined through a composite factor, and then raked altogether.