

Attachment L

Description of Power and Precision

Attachment L

DESCRIPTION OF POWER AND PRECISION

The present document describes the findings of a power analysis that was performed to determine the extent to which statistically significant differences could be detected when comparing the results of survey data collected in 1999 by KFF to data collected in 2008 and 2010 by OMH/NORC and to be collected again in 2012. First, a brief description of the concepts being measured and the methods used to compute them are described below.

Description of the Concepts Being Measured and Compared

The primary objective of the proposed data collection is to measure the relative change in the US public’s awareness of racial and ethnic health disparities between 1999 and 2015. A survey fielded in 1999 by Kaiser Family Foundation and Princeton Research Associates International (KFF/PSRAI) reported that approximately 47.8 percent of a sample of randomly selected respondents indicated being “aware” that certain racial and ethnic health disparities existed. This “awareness” measure was calculated as a percentage of the respondents who reported that they felt that the reference group (African-Americans or Hispanics) were disparate in terms of the health outcomes achieved. Specifically, the percentage was derived from the responses to six sub-questions (from four core questions) included in the 1999 KFF/PSRAI survey instrument (see Exhibit 1 below). These questionnaire items ask respondents to indicate whether they perceive African Americans and Hispanics as faring “better off,” “worse off,” or “the same as” the average White (non Hispanic) American in various health outcomes where research has shown (and continues to show) that significant disparities exist.

Exhibit 1: Survey Items Used to Compute Awareness of Health Disparities

Questionnaire Item (2008)	Response Options	Item Number (1999 Survey)
1. Do you think the average <i>African American</i> is better off, worse off or just as well off as the average <i>white person</i> in terms of: b. infant mortality—that is, a baby’s chance of surviving after birth; c. life expectancy—that is, how long someone can expect to live?	1. BETTER OFF 2. WORSE OFF 3. JUST AS WELL OFF 8. (DON’T KNOW) 9. (REFUSED)	Q11b-c
4. Do you think the average <i>African American</i> is better off, worse off or just as well off as the average <i>white person</i> in terms of: d. having health insurance?	1. BETTER OFF 2. WORSE OFF 3. JUST AS WELL OFF 8. (DON’T KNOW) 9. (REFUSED)	Q15d
6. Do you think the average <i>Hispanic/Latino</i> is better off, worse off or just as well off as the average <i>white person</i> in terms of: b. infant mortality—that is, a baby’s chance of surviving after birth; c. life expectancy—that is, how long someone can expect to live?	1. BETTER OFF 2. WORSE OFF 3. JUST AS WELL OFF 8. (DON’T KNOW) 9. (REFUSED)	Q16b-c
10. Do you think the average <i>Hispanic/Latino</i> is better off, worse off or just as well off as the average <i>white person</i> in terms of: d. having health insurance?	1. BETTER OFF 2. WORSE OFF 3. JUST AS WELL OFF 8. (DON’T KNOW) 9. (REFUSED)	Q20d

It is important to note that the core questions that were included in the 1999 survey were unaltered and included in the 2008 and 2012 versions of the instrument so that comparisons could be made without introducing additional error to the computations. The research team has also taken special care not only to replicate the items, but to place them in the relatively same order as they appeared in the 1999 survey

instrument. There are only small differences in terms of the placement of the core items, the 2008 and 2012 questionnaire has the core items appearing slightly earlier than in 1999, but are not so far apart that would introduce any type of error or question order effects.

Method of Computing “Awareness” of Health Disparities

To compute the percentage of respondents that are aware of health disparities involves analyzing the responses to six questions that speak to the respondent’s knowledge of health disparities for a number of health outcomes and conditions. A first step in computing awareness requires that response options for the six core items “better off” and “worse off” are recoded to “1” (indicating that the respondent is aware of a specific disparity) and the “same as” responses to “0” (indicating that the respondent does not perceive a disparity). Next, an indicator variable is created which specifies a “1” if the respondent indicated that they perceived a disparity for any of the items focusing on specific disparities, and “0” for respondents that did not indicate that a disparity was present for any of the items. It should be noted that this method of creating a 0-1 indicator variable produces a variable for which the underlying distribution is binomial and the resulting estimates of the proportion of respondents coded “1” (i.e. those aware of more than one disparity in at least one subgroup) may be assessed in light of that distribution.

Rationale for Selecting the Sample Size and Expected Power

The 1999 study had a sample size of 3,884 with a design effect (DEFF)¹ of 2.11. If the exact sample size and design effect were replicated for the 2008 survey, the power to detect a difference of at least two (2) percentage points would be approximately 29 percent. Given that the 1999 sample cannot be modified (i.e., increased), the research team considered increasing the sample size for the 2008 data collection and found that increasing the sample has limited effects on the power to detect differences (see Exhibit 2 below).

Exhibit 2: Sample Size and Power Required to Detect Two Percentage Point Difference Between 1999 and 2008

Sample Size	Expected Power*	Increase in Power	Required Sample Increase
3,884	29	0	0
4,460	57	+05	1,116
7,500	64	+12	3,616
10,000	68	+16	6,116
15,340	72	+20	11,456
139,900	80	+28	136,016

*Assumes a design effect equal to, or smaller than, the one observed in 1999.

Exhibit 2 shows that increasing the sample yields modest increases in power that plateaus at about 72 percent. Thus, increasing the sample size above 15,000 would not yield significant increases in power

¹ Refers to the required magnitude of the observed difference in order for it to be considered significant.

until the sample size is increased well above 100,000. A key consideration is that the power level available to the ongoing assessment of these data is constrained by the fact that the researcher has no control over the 1999 sample size. That is, the 1999 sample size is fixed, so the power of a test comparing new data to the 1999 observations can only be influenced by the sample size, albeit with reasonable limits. The increase in power reaches a plateau in the low 70 percent range such that increases in sample size beyond approximately 4,500 do not have an appreciable influence on power. Since the comparison needs to be made to the 1999 data, the question becomes one of realistic sample sizes and the power they can achieve. The fact is that the power cannot be increased even to a conventional level of 80 percent without extremely large sample sizes. Nonetheless, our experience has been that power levels this size (and even lower) are not uncommon in trend studies where the sample for the first year of data collection is fixed. Should the effect size increase in the 2008 data collection, the research team will have the ability to detect a significant differences the majority of the time.

Based on all the considerations described above, the research team concluded that that the most cost-effective and efficient sample size would be a sample of the same size as 1999 up to 4,460 completed interviews. This would achieve a reasonable increase in power over the 1999 sample without undue burden on respondents, as well as excessive expense, time, and resources to collect the data.