Durable Nursery Products Exposure Survey (DNPES) Response to OMB Comments

1. Supporting Statement (SS) Part A, page 1 – please be careful not to use words like "impact" when there is no experimental design underlying the analysis.

   Response:

   We have made this change as part of the resubmission.

2. SS, item A9 – The standard rate for federal cognitive labs is $40 (not $50) for participants in their cog labs. Please update accordingly.

   Response:

   We have made this change as part of the resubmission.

3. Regarding changes to the questionnaires (as a result of the cognitive labs or other changes), please note that this will require re-submission to OMB. We have an expedited review process for changes that are "non-substantive" – and we would assume that changes to questionnaires due to the cognitive labs will fall into that bucket. If, however, the questionnaires change dramatically (*e.g.*, different respondent, entirely different subject matter), CPSC may need to seek additional public comment before proceeding.

   Response:

   We have resubmitted the revised questionnaire, along with the supporting statements, as requested. The changes were "non-substantive." Should additional revision be required due to cognitive testing, we will resubmit the questionnaire. Any changes will be "non-substantive."

4. SS A16 – Please provide additional detail in the analysis section. We would like to see a more developed discussion of how the survey data will be compared with the injury data, and what estimates will be created. In addition, how will CPSC develop valid estimates of intensity of usage (specifically mentioned) given limitations of the sample design? Finally, please describe (in more detail) modeling planned. In each case, we are especially interested in clarity about the unit(s) of analysis how CPSC knows whether the current sample design and size is sufficient to support it.

   Response:

   The goal of the analysis is to estimate the prevalence of, and exposure to, 24 different infant products in U.S. households. These estimates will be used in conjunction with other CPSC injury data to estimate injury rates for the products.

In general, the unit of analysis is the household, and the target population is households with children age 5 years and under.

Analysis: Simple Risk (or Injury Rate) Comparison

As an example illustrating how the data will be used, consider the case of infant bath seats. One long-standing policy issue for the agency has been whether an infant death is more likely to occur with the use of an infant bath seat versus the use of only a bath tub. If an infant death is more likely to occur with the use of a bath tub only, then efforts to ban infant bath seats or reduce their use could have the undesired consequence of increasing the risk of infant bathing deaths, which would be the exact opposite of the CPSC's policy goals. The rate of death associated with these products would be compared on a per-child basis:

$$\frac{\text{number of bathtub only-related deaths}}{\text{number of children using bathtubs only}} \; vs. \; \frac{\text{number of bathseat deaths}}{\text{number of children using bathseats}}$$

For the numerators, the agency compiles death data from a variety of sources. However, the agency lacks the information needed to estimate the denominators in the above calculation, which is the number of children using only bath tubs versus the number using bath seats. One purpose of the DNPES is to collect the data needed to conduct these rate comparisons, not only for bath seats, but also for other products.

Analysis: More Formal Modeling Efforts

In addition to the simple injury rate calculations described above, the survey results could be used to conduct a formal logistic analysis of the factors associated with nursery product risks. Such an analysis could be done if the exposure survey data can be compared to parallel injury data that might be collected from a survey of injuries reported through the National Electronic Injury Surveillance System (NEISS). In order to conduct this type of analysis, we would need to compare the characteristics of injured victims (from the parallel injury survey) with the characteristics of non-injured users (from the exposure survey). Consequently, to do this type of analysis, we must be able to exclude cases from the exposure survey in which children were injured. That is why we want to collect information on whether the children using nursery products were injured. We do not intend to use the information on injuries from the exposure survey to evaluate injury patterns—rather, we would use information from NEISS to do that.

The impetus for asking the injury questions in the exposure survey is based on lessons the agency learned from conducting an ATV exposure survey in the 1980s. In that study, data on the ATV users who were injured were not collected in the exposure survey, which made the formal analysis of risk patterns more difficult.

Later, in the 1990s, the agency conducted a more successful survey in which injury data were collected. The later ATV study provided sufficient data to support a formal modeling effort. Besides some simple risk comparisons, we envision applying the methods of the ATV study to the DNPES data. Ultimately, the specifics of such an analysis would be tailored to the product of interest and will vary across the different products for which data are collected.

In the later ATV study, the CPSC was able to model formally the factors associated with risk by comparing injured ATV riders to uninjured riders, as a function of user characteristics, product characteristics, and use patterns (see Attachment A). Similarly, the agency conducted an analysis of bicycle risk patterns in 1994 (see Attachment B). One key to being able to conduct this type of analysis is determining which users from the exposure survey were injured, as described above. From a policy standpoint, this type of analysis has proven useful to the agency in the past because it provides information on the types of situations where injury is more likely to occur, which, in turn, allows us to target regulatory efforts, as well as informational and educational campaigns, more efficiently.

Intensity of Usage

"Intensity of usage" is not a concrete term that can always be collapsed into a single measure or estimate. Rather, it is a descriptive term that embodies several dimensions of product use, and the relevant dimensions will vary according to product. Generally, intensity of use will be captured by the responses to several questions in the survey, questions that have been tailored to each specific product. These questions cover such dimensions of use intensity as: the number of particular products in use in the household, the frequency of use, and, in some cases, the duration of use. The intent of these questions is to provide information on household exposure to product risk, information beyond whether a household simply owns a product.

As a specific example, consider the case of high chairs. To capture the dimension of use intensity, the survey asks how many high chairs a household owns and how many they use. Additionally, for the most frequently used high chair, the survey asks how frequently the high chair is used (for example, every day) and the duration of use (for example, 30 minutes). From the information collected, it will be straightforward to compute "intensity of usage" as "minutes in use per day," a measure that can be compared across households, as well as for other products where "minutes in use per day" can be computed similarly. Obviously, such a measure captures household exposure to a product-related risk more meaningfully than a simple question of whether the household owns a product.

Sample Design and Size

The CPSC has undertaken measures to ensure that the current sample design and size will be sufficient to support the envisioned analyses. Using available data (from American Baby Group and elsewhere), we have made preliminary estimates of the prevalence of ownership of each product in the general population. While these preliminary estimates are not based on a probability-based sample, they are useful for selecting the product modules each respondent will be asked to complete. These selection rates will be updated, however, as the survey is rolled out. Therefore, if we initially believe that few respondents will be eligible for one module but end up with numerous completes, we will shift priorities to ask respondents about products for which we have fewer completes first. We plan to make these updates on an ongoing basis. The product selection rates used for each household will be stored for later use in weighting responses to produce national estimates.

In this manner, we will target a minimum of 250 completes for each product module. This means that products used less often will be selected at a much higher rate than more commonly used products. This should allow for making estimates about a particular product with adequate precision, such as, for example, the number of baby sling users reporting a death or injury. For an estimated 50 percent proportion, a sample size of 250 would give a 95 percent CI half-width of 6.3 percent, assuming a design effect of 1 (*i.e.,* no variation in household weights).

Of course, there may be products for which there are fewer eligible respondents. For example, except for a few products, respondents are eligible only if they use/used the product at least a few times a week. Therefore, it is possible that we could end up with fewer respondents in some cases and, in the extreme case, too few respondents for a particular product module to do any modeling using the product-specific questions in the product module. However, the prevalence can be estimated for all product modules because the precision for each prevalence rate is based on the number of households completing the inventory portion of the questionnaire (*i.e.*, the section that asks whether the household uses or has used each of the 24 products). This is expected to be about 1,800 households (16,667 completed screeners * 0.132 eligibility rate * 0.80 extended questionnaire response rate = 1,760).

5. Please provide more information about the planned mode effects test.

   Response:

   Mixed-mode survey design introduces the possibility for mode differences, because respondents may answer a question differently, depending upon the mode in which it is asked. Mode differences can have serious implications for data

comparability.[1] A question presented orally in a telephone survey may prompt a very different response than the same question presented visually on a Web or paper survey. Likewise, a question may be interpreted differently when it is presented within the context of a paper questionnaire than when a respondent sees or hears only a single question at a time in a Web or telephone survey (a phenomenon known as "segmentation"). There are many other variables that may trigger mode differences, including: dynamic (Web, phone) versus static (paper) presentation; oral (phone) versus written (paper) versus typed (web) transmission; and interviewer administration versus self-administration.[2]

The literature suggests three general approaches to designing multi-mode surveys to minimize measurement error. However, no consensus exists on which is the best approach.[3]

In the first approach, mode-enhancement, one mode is considered to be the "primary" mode. Therefore, the goal is to design the best possible instrument for that single mode, concentrating on obtaining the highest quality data—even if that requires sacrificing equivalency across modes. This approach is generally advocated when the secondary modes are used only sparingly to increase response rates or coverage.

The second approach, advocated by Dillman, is the unimode design. With this approach, questions are written and presented identically across all modes. This requires designing questions that are suitable for administration in all modes, which one researcher has called a "one-questionnaire-fits-all design."[4] Dillman outlines guiding principles to ensure that questions are effective, regardless of mode, including reducing the number of response categories to make questions appropriate for both visual and aural presentation. As a result, however, formats that are less than optimal for a mode may be used.[5]

The third approach has been called generalized, universal, or mode-specific design. This approach contends that respondents process information very differently in different modes. Therefore, the same question may have a different meaning in different modes. Paradoxically, to achieve comparability across modes, it may be necessary to change question formats.[6] Rather than presenting

---

[1] Roberts, C. (2007). Mixing modes of data collection in surveys: A methodological review.

[2] Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). Internet, mail, and mixed-mode surveys: The tailored design method. Hoboken, New Jersey: John Wiley & Sons: 326-329.
Pierzchala, M., Wright, D., Wilson, C., & Guerino, P. (2004). "Instrument Design for a Blaise Multimode Web, CATI, and Paper Survey." In Proceedings of the 9th International Blaise Users Conference 2004. Gatineau, Canada: Statistics Canada, 2004 (with Debra Wright, Claire Wilson, and Paul Guerino).

[3] Dillman 2009, Roberts 2007, and Pierzchala, M. "Disparate Modes and Their Effect on Instrument Design." In Proceedings of the 10th International Blaise Users Conference 2006. Arnhem, The Netherlands: Statistics Netherlands, 2006.

[4] de Leeuw, E.D. To mix or not to mix data collection modes in surveys. Journal of Official Statistics-Stockholm- 21 (2):233, 2005.

[5] Dillman 2009.

[6] Pierzchala et al. 2004.

identical questions across modes, this approach seeks "cognitive equivalence" across modes. E.D. de Leeuw compares this approach to "modern theories on questionnaire translation, in which not the literal translation, but the translation of concepts is the key."[7]

It is not always clear, however, what the best way is to obtain this "cognitive equivalence." Frequently, multi-mode survey design is performed on an ad hoc basis. A survey may be designed for one mode and subsequently adapted to another. Conversely, survey design across modes may happen concurrently but with little coordination between modes. Pierzchala has argued that many survey researchers "often overlook the importance of a holistic perspective."[8] Generally, this is due to cost or time limitations because multi-mode design is often not straightforward. "Adjustments to each mode are likely in order to achieve 'cognitive equivalence' between the modes in the mind of the respondent. These adjustments may include changes in question text, data definition, routing, fills, and so forth. Thus it may be necessary to leave time for prototyping and assessment, and reworking of the instrument."[9] Such constant reworking and revision is not always a practical option.

For the most part, the DNPES uses the unimode design. Some examples of the ways in which we minimized differences across the telephone and Web modes include:

- The questionnaire originally contained numerous open-ended questions along with a list of unread response categories for telephone interviewers to use for coding respondents' answers. All of those items were converted to closed-ended items with the response options read to the respondent (along with an "other specify" option). The purpose was to present the same preset response categories to telephone respondents orally as Web respondents received visually.
- Where possible, kept the number of read response options to five or less.
- Adopted pronouns that worked in both modes (*e.g.,* "you" instead of "I").
- Streamlined the product definitions for verbal delivery (the same definitions will be presented to Web respondents).
- Used wording and placement to minimize social desirability bias in the telephone administration of the co-sleeping items, enumeration items that ask for child's name, questions about leaving the child alone in the product, and the premature birth/disability items.
- Wrote section transitions for the telephone instrument that will also be displayed in the Web instrument.
- Plan to mail product photos to telephone respondents so that they have access to the same visual aids as the Web respondents.

---

[7] de Leeuw 2005.

[8] Pierzchala, M. "Experiences with Multimode Surveys." In Proceedings of Statistics Canada Symposium 2008. Ottawa: Statistics Canada, 2008.

[9] Pierzchala et al. 2004.

6. Please provide details of the planned precision analysis and also describe how the estimates provided will meet program goals.

    <u>Response</u>:

    Estimates of product incidence rates are available from the American Baby survey. While these results seem to suggest that even the lowest incidence products should have sufficient samples for reasonable precision, these results seem to be from a population of households with very young children (maybe through 18 months) rather than through 5 years (our planned population), so some of these items may be less prevalent in our sample. In short, it is very hard to project this because this is one of the key pieces of information this survey is designed to discover. We will be monitoring product prevalence to adjust the probability of module selection for each product throughout the survey. If we find any products where overall completion numbers are projected to be too small for meaningful precision, we may reconsider a Web panel for which we would seek separate approval, as discussed in item No. 11.

    The sample sizes should allow estimation of the prevalence rate for each product and a measure of exposure, as discussed in item No. 4.

7. Including a Spanish language component to the survey is useful. However, Census and other literature gives some specific findings that suggest the importance of targeting households and also of the structure of the mailing piece (e.g., 2 questionnaires versus a 2-sided one). Please explain how CPSC can adapt the current plan to more closely mirror what the literature suggests is a best practice.

    <u>Response</u>:

    We are unaware of much literature on questionnaire design for targeting Spanish-speaking households, particularly where the language spoken in the household is unknown. There is an unpublished 1994 Census study by Corteville that discusses targeting mailers to Spanish households. Corteville found that including a Spanish language questionnaire, in addition to an English version, significantly improved the completion rate by 2 to 6 percentage points targeted in the strata with housing units in Census blocks with 30 percent or more Spanish-speaking housing units, with no difference in the strata of 15-30 percent or more Spanish-speaking units.

    The same unpublished Census study also considers two different Spanish language implementations—two forms versus one bilingual form—but found that neither implementation significantly outperformed the other. The Census ACS uses a "swimlane" design (English and Spanish side-by-side), but there are no known publications on the performance of this design. One of the NHES field

tests found that the response rates for separate questionnaires were similar to those of a bilingual "swimlane" version.

Based on this information, we propose using an English-only mailing for the first round and using two-sided (English on one side and Spanish on the other, where the respondent would flip the page to see the Spanish version) in any subsequent mailings. This approach would also partially address concerns about respondents using the Spanish portion of the screener to fill out the English portion (see No. 17d).

8. What is Westat's past experience with mailing pictures in advance of a telephone interview that requires their use? Why is this approach being proposed?

   Response:

   While the interview does not necessarily "require" the use of these pictures, we found in preliminary cognitive testing that certain products were confusing when described but were recognized immediately when pictures were provided (for example, bouncers were often confused with jumpers). In other cases, pictures can confuse respondents, as was the case with bedrails (it was not always clear to respondents that the beds were adult-sized). Therefore, pictures can help respondents understand the specific type of product the questionnaire is asking about and reduce bias that might arise from inconsistent definitions or understanding of the products covered by the survey. Also, because we plan to make a Web version available to respondents, which would include the pictures, we wanted to assure that phone respondents would have access to the same information when interviewed. This is an example of how we have tried to minimize mode effects in the development of the survey (see item No. 5). Westat has used illustrations in past successful surveys; for example, it has used illustrations to demonstrate portion size for food and nutrition work.

9. SS Part B, item 2.4 says that the feasibility of oversampling for low-income households will be investigated. Has this already been done? What are the conclusions? How has the sample design and size changed as a result, if any? How will this affect the precision analysis?

   Response:

   As with the Web panel (see item No. 11), this option was under consideration when funding levels were lower. With the increased funding and the resulting increase in the current planned overall sample size, we expect to get a sufficient number of low-income households included (even if these households respond at a disproportionately low rate). The 2010 Current Population Survey (CPS) data indicates that 23.2 percent of households with children age 0 to 5 are below 100 percent of the federal poverty level and 44.2 percent are below 200 percent of the poverty level. Based on this, we would expect about 408 responding, eligible

households below 100 percent of the poverty level, and about 778 responding, eligible households below 200 percent of the poverty level. These sample sizes are adequate for producing estimates for low-income households without oversampling them.[10] Although the address-based sampling (ABS) frame could be stratified on the percentage of low-income households in the county (using 5-Year 2005-2009 ACS data) and addresses in the low-income stratum could be oversampled, this would increase the variability of the household weights. If households could be stratified directly on a low-income indicator, oversampling them to target 1,000 below 100 percent poverty (out of a 2,000 total sample size) would result in a design effect of 1.32. Given the high poverty rates among households with children and the large overall sample size for the DNPES, it does not seem worthwhile to oversample low-income households. This result is very consistent with the findings by Waksberg, Judkins, and Massey in their Survey Methodology article in 1997 for oversampling in face-to-face surveys.[11]

The supporting statements have been updated to take this information into account.

10. SS Part B, item 2.4. As a result of the item above, is an income question being proposed as an addition to the screener? Note that this question may have an effect on response rates (this should be investigated) but also may be useful for nonresponse bias analysis.

Response:

We would like to avoid asking about income in the screener because it might reduce response rates and, as noted in item No. 9 above, we are not proposing oversampling by income. Instead, income questions are asked at the end of the survey where, if the question is perceived negatively and the respondent drops out, the majority of the survey will have already been completed. Instead, we could ask less intrusive questions that could serve as a proxy for income, such as own/rent home (preferred) or length of time in home. (See also responses to items No. 12 and No. 17a.) Our limited budget precludes conducting both an investigation of the impact of using income questions (or alternative additional questions) on the screener and the actual survey.

We are considering adding additional questions to the screener that would help estimate the nature of nonresponse at the second stage. Please see the responses to items No. 12, No. 16, and No. 17a for additional information on this issue.

---

[10] We recognize that low-SES household's often have lower response rates. However, we still anticipate having a sufficient sample size for meaningful analyses.

[11] Waksberg, Judkins, and Massey, "Geographic-Based Oversampling in Demographic Surveys of the United States," *Survey Methodology* Vol. 23, Number 1 (June 1997).

11. Is CPSC seeking PRA approval now for the potential supplemental sample via a web panel? If so, we need to understand more about how that will be implemented.

    Response:

    No, CPSC is not seeking approval for a potential supplemental sample. This option was considered as a possibility when funding levels were low enough to significantly limit the number of completes we could expect for any given product (see also item No. 9). This option was added to allow for the possibility that product-specific information for some more rare products might need to be supplemented. Since that time, however, additional funding has become available to allow us to increase the number of respondents we can survey, making supplemental panels unnecessary.

    However, if other factors arise that reduce the number of completed surveys, it is possible that a web panel might again become a desirable option. Should this occur, the CPSC will take any actions necessary to comply with the Paperwork Reduction Act.

    The relevant sections of the supporting statements have been updated accordingly.

12. While the "nonresponse follow up subsample" plan should be useful in comparing something akin to what in the literature would be called "early" and "late" respondents, it is not clear that this gives much information on true nonrespondents. Given the anticipated screener response rate, we think this should be more robust.

    Response:

    We recognize that even surveys with very high response rates and large sample sizes can have non-ignorable nonresponse bias. Efforts to increase response rates to anything short of 100 percent would still leave us without an understanding of nonrespondents. The plan to study nonresponse bias involves four strategies: (1) a nonresponse follow-up subsample to compare respondents who respond with additional contact effort to those who respond without the extra effort; (2) response propensity analysis as part of the nonresponse adjustments in the weighting; (3) under-coverage analysis as part of the poststratification in the weighting; and (4) comparison of DNPES estimates with external sources, where possible. The nonresponse follow-up subsample is discussed in section B.3.4.2 of the original OMB submission. Response propensity analysis and poststratification are discussed in more detail below.

    A nonresponse analysis will be done to profile screener nonrespondents as part of the nonresponse adjustments in the weighting. This involves evaluating which ABS frame variables (among those with a small percentage of missing data),

Census and 5-Year 2005-2009 ACS demographic variables are most effective in distinguishing between subgroups with different propensities to respond. Census data can be obtained for the census tract or block group containing the sampled addresses, because Census geography codes are provided on the ABS sample records. The 2005-2009 ACS data can be obtained at the county level. We plan to use CHAID software for this purpose. CHAID is a commonly used tree-based algorithm for analyzing the relationship between a dependent variable and a set of predictor variables. The software forms cells of households with similar response propensities, using household characteristics that are available for all sampled households. A nonresponse adjustment factor equal to the inverse of the weighted response rate for each cell is then applied to the household base weight. This adjustment restores the distribution of the household respondents to match that of the original sample. Response rates will also be calculated by the variables identified as correlates with response propensity by the CHAID software. This will enable us to develop a profile of the screener nonrespondents. To the extent these correlates are known to be also correlated with the survey response variables, this may give an indication of the direction of the nonresponse bias.

The nonresponse-adjusted household screener weights will be poststratified to state totals of households with children under age 6, at a minimum. These control totals can be constructed by multiplying the household totals at the county level and the percent of households with children ages 0 to 5 from the 5-year 2005-2009 ACS at the county level, then summing to the state level. Note that the 2005-2009 ACS does have the number of households with children under age 6 below 100 percent of poverty levels at the county level. However, because income is not being collected in the screener (and a single item or even two on income would not be sufficient for creating poverty income levels in a way that is reliable), we cannot include low-income as a post-stratification dimension. But if a variable correlated with income, such as tenure (whether the person rents or owns) is included in the screener, and the item response rate is relatively high, tenure could be included as a post-stratification dimension and, in this way, we could try to improve estimates of low-income households. The post-stratification factors will indicate demographic groups (such as renters) or geographic areas of under-coverage in the screener respondents (though there will be some sampling error in these adjustment factors). Also, the Census Bureau has released 2010 Census population and household totals by age/race/sex for all states down to the block level. Because the Census county, tract, and block group codes will be provided on the ABS sample by the vendor (and can also be geocoded in-house at Westat from the addresses), the Census totals can be compared with the distribution of the weighted DNPES household sample by geography and demographic subgroups to assess under-coverage.

The supporting statements have been updated accordingly.

a. What other strategies did CPSC and Westat consider and what else might they propose to address this concern, especially for first stage (screener)

nonrespondents? We are looking for a fairly significant proposal here, as the projected response rates fall far below what would be required for a survey of this importance and intended use.

Response:

The response propensity analysis and poststratification steps in the weighting will be the primary tool for understanding non-respondents (see above). This has been part of the plan from the beginning, but it was not discussed thoroughly in the Supporting Statements. (See also our responses to items No.16 and No. 17.)

b. One strategy for measuring nonresponse at the questionnaire stage is the inclusion of one or two substantive (rather than merely screening) questions on the screener. What are the one to three substantive variables (rather than demographic variables) that CPSC believes could be used to have measure second stage nonresponse?

Response:

See response to item No. 10 above, as well as item No. 17a.

13. Please provide the screener questionnaire formatted as you would for mailing (e.g., Dillman use of colors, segmentation, etc). We will review is as formatted and make specific recommendations then.

Response:

We will base this on the most effective NHES format. Prior to laying out the screener and providing it to OMB formally for review, we would like to get your feedback on the screener content items discussed here. This will allow us to use our limited budget more effectively.

14. On the Concluding questions section of the main questionnaire, the race question instruction should be "please select one or more." In addition, in line with OMB standards, CPSC may not include the category of "other."

Response:

This has been modified in the resubmitted questionnaire.

15. Please explain in some detail why CPSC is asking about a chronic health condition or disability. Please note that the literature is clear that it is highly unreliable to ask this as a yes/no question. Depending on your answer, we will help you to identify a more appropriate question/set of questions to use, as this is an area where we have been working with agencies to standardize measures.

Chronic health and disability questions are asked because the use guidelines for durable nursery products (*e.g.,* the age) may not be applicable under these conditions. For example, children born prematurely can use bassinets well beyond the recommended age limit of 5 months. Similarly, a child with a developmental disability who is unable to push up on their hands and knees at around 5 months of age could safely use a bassinet until they could push up on their hands and knees.

Frequently, nursery products are used differently for children with health conditions and/or disabilities. Therefore, we would want to exclude them from some of our analyses or, at least, consider them separately. For example, when analyzing the start- and stop-use questions, we would want to exclude cases where the children were born prematurely or had a chronic health or disability issue that might render the usual guidelines for product usage meaningless. Asking for additional details about the condition or disability will allow our human factors experts to determine where data should be left out and where it should be left in for these analyses.

16. We are concerned about the expected response rates. While we appreciate realistic estimates we think that when the expected response rates are very low (as here), a lot of effort should be placed on nonresponse bias analysis. We also think that more can be done with the screener to get response rates up (NHES, for example, is over 50%). Even with higher response rates, however, we would expect to see more nonresponse bias analysis. Our main concern is that without more attention to these areas, CPSC will find itself with the same problem that American Baby has – no poor households, etc.

    Response:

    In response to this portion of OMB's additional comments, we would like to summarize several points and identify additional discussions, where applicable:

    - We believe our response rate estimates are conservative. Because the target respondents (parents of very small children) are likely to be more interested in participating in this survey about the safety of children's products, response rates may be higher than we assume. Early tracking of response rates will help provide better estimates, which will be used to adjust the released sample size.
    - As discussed in item No. 12, we always intended to use response propensity analysis and poststratification in the weighting to assess the potential for nonresponse bias and under-coverage. (See also response to No. 10.)
    - We are considering additional questions for the screener but wish to limit the number of added variables because of the potential negative

consequences and the fact that our budget will not allow us to conduct an investigation of their impact prior to conducting the survey itself (see items No. 10 and No. 17 for more).

- We believe that our ongoing product prioritization strategy (see the last section of item No. 4 ("Sample Design and Size") and item No. 6), combined with our approach to nonresponse bias (see item No. 12), should address concerns about the survey results containing no low-income households.

17. After talking to the program manager for the National Household Education Survey (NHES) last week about the latest field test results, here is what that program is finding to be the very best screener/mailing package/Spanish mailing strategy. Use of this strategy should improve the anticipated response rates a good deal for the CPSC survey (NHES obtained more than double what CPSC is estimating).

  a. A brief screener that does not ask for names or other personally identifiable information (PII) formatted in the Dillman style. See attached. The prominence in the CPSC questionnaire of the adult name, phone number and email in the draft we say, and the dearth of other questions seems likely to be a turn off. It may be useful to highlight the one question about presence of children, and considering adding perhaps one item to it that would be useful for nonresponse analysis, then deemphasizing visually the PII.

  Response:

  We are concerned that the description of the very best approach from the NHES has not been evaluated fully. Some of the results from the screening phase may suggest good performance, but they must be evaluated fully with respect to the overall response rates. The Westat staff working on this analysis has not yet completed that work. As suggested, we will highlight the question about the presence of children and visually deemphasize the PII questions when we submit the screener layout to OMB (see response to item No. 13). We note that NHES did not have an income question on the screener, and we propose to avoid that too. We also have concerns about adding unnecessary questions (*i.e.,* unasked elsewhere in the survey) merely as a proxy for income, even if it would potentially allow for additional nonresponse bias analysis (again, see item No. 10).

  We are considering the idea of moving questions regarding composition of the household (age and sex of household members only) to the screener. This would seem to address your concerns about asking PII in a screener where little else is asked, as well as minimize the burden on respondents (*i.e.,* not adding additional questions to the survey overall). This would

also allow us to reconfirm in the extended survey that the children under the age of 6 live in the household before we get their name/nickname information. .

We do not have (and cannot get) the funding that would be necessary to test two alternative screeners adequately. We would appreciate additional feedback/discussion prior to laying out a template screener.

b. A $2 prepaid cash incentive in the screener mailing package (note that the NHES $5 incentive performed best but the extra increase of about 4.5 % does not appear cost effective compared to the $2 incentive).

Response:

While we appreciate how incentives for the screener mailing package could boost response rates, it is simply not in our budget, given the large number of mail-outs that are planned.

c. Sending a Spanish questionnaire along with an English questionnaire ONLY at the second and third mailings to all addresses (note that this performed better than a universal first mailing and better than targeted mailings by geography and by surname)

Response:

As discussed in item No. 7, we will incorporate this suggestion.

d. A Spanish questionnaire design that makes it intentionally difficult to use the Spanish portion to complete the English portion. NHES found that earlier designs appeared to be encouraging Spanish speakers to complete the English version (with the Spanish "column" as a language guide) which then triggered receipt of the much more difficult to complete topical questionnaire in English. These households then couldn't successfully complete the topical questionnaire in English. Therefore, NHES recommends either use of two separate forms in the same mailing package or two-sided with all English on one side and all Spanish on the other, as both performed approximately equally in the NHES test.

Response:

As suggested (and discussed in item No. 7), we will include a two-sided English/Spanish questionnaire for additional mailings. Additionally, we will modify the Spanish version of the screener to make it difficult to use it as a guide to answering the English version of the screener.

e. Prenotice letter only if funding permits (provides early gains but effect wore off after third mailing).

Response:

The design of this survey includes a number of contacts with potential respondents. While they are not literally pre-notice letters, they have the effect of a prenotification. The mail screener includes a cover letter with information about the CPSC and the survey. Once again, the effect of the experiment in NHES on the use of a separate prenotice letter is being studied, but initial analysis shows little or no gain in response rates.

f. Fedex for third mailing only (boost of about 6% over USPS priority mailing; costs about $2 per case).

Response:

We will consider adding this option should additional funding become available.

g. The NHES nonresponse bias strategy includes assessing frame variables from vendors, and is admittedly still under review. However, the frame variables appear to be useful in identifying households with children and some basic demographics.

Response:

Westat statisticians involved with the NHES survey tell us that the NHES: 2011 Field Test data are still under review; however, the NHES: 2009 Pilot Study data have shown that the presence of children in the household variable on the ABS frame is not useful for identifying households with children, due to a high rate of missing data and low correlation with the self-reported status when data are available for comparison. If the NHES: 2011 Field Test review indicates otherwise, we will use this and other demographic variables with low missing rates in conjunction with Census data in the nonresponse bias analysis.