Supporting Statement B For:

Durable Nursery Products Exposure Survey

(CPSC)

May 30, 2012

Jill L. Jenkins, Ph.D., CPSC Project Officer
Economist
Consumer Product Safety Commission
Directorate for Economic Analysis
U.S. Consumer Product Safety Commission

U.S. Consumer Product Safety Commission
4330 East West Highway
Bethesda, MD 20814

Telephone: 301-504-6795
Fax: 301-504-0124 and 301-504-0025
JJenkins@cpsc.gov

TABLE OF CONTENTS

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

B.1 Respondent Universe and Sampling Methods

Different approaches will be used to select respondents at the two different stages of the DNPES. The first stage of the survey - cognitive interviews - will use purposive samples in order to efficiently test the questionnaire with specific subgroups of interest. The second stage of the survey – a large nationwide field survey - will employ statistical sampling methodology to enable projections to the population, non-response analysis, and a mode effects evaluation. This section provides a description of the respondent sampling and data collection procedures to be used for each stage of the survey. Section B.1.1 describes the sampling procedures that will be used in the cognitive interviews, and Section B.1.2 describes the procedures that will be used in the main field survey. Non-response analysis is addressed in section B.3.4 and mode effects are covered in section B.3.5.

B.1.1 Cognitive Interviews

For the cognitive interview, CSPC has identified subpopulations of particular interest to represent diverse racial/ethnic groups, primary language, and socioeconomic status. Purposive sampling is designed to efficiently reach and recruit parents or guardians of young children from these key subpopulations and ensure that Spanish interviews are conducted. It is of particular importance to interview people of low socioeconomic status, whose level of education can be a barrier to comprehending and following the survey instrument, and persons not fluent in English, for whom innovative ways of communicating survey questions may be necessary.

This study will include up to 18 cognitive interviews in English and up to 18 cognitive interviews in Spanish. For English cognitive interviews, respondent recruitment efforts will ensure that key racial and ethnic groups (non-Hispanic/White, African-American, and Hispanic) are represented with a goal of at least 4 cognitive interviews with people from each group. Additionally, recruitment efforts for the English group will ensure that at least 3 cognitive interview respondents are from low socioeconomic status (SES). Recruitment for Spanish cognitive interviews will target 5-6 Spanish-speakers who are of low SES.

The process for recruiting low socio-economic status (SES) participants for the Spanish-language cognitive interviews involves Spanish-speaking recruiters and recruiting materials at locations with a likelihood of contacting the population. Community centers, social service agencies, educational

facilities where English-as-a-second language classes are offered, or other forms of social contact have been used a sources for identifying the population. We have used both in-house and outside recruiters to contact potential participants and conduct the screening interviews to determine eligibility for the study. Some Spanish-speakers may be recruited by phone using phone numbers provided from commercial vendors that are known or are highly likely to have Spanish-speakers in the target income range. Before being asked to participate in the cognitive testing, all potential participants complete a brief screener survey in Spanish that confirms that their primary language is Spanish and also collects their household income level.

B.1.2 Nationwide Field Study

The DNPES target population is households with children younger than 6 years old in the civilian non-institutionalized population of the United States. The field survey will be conducted nationwide using a representative and probabilistic address-based sample to collect information about the ownership and usage of various types of children's products. The sampling frame will be constructed from a commercially available list of mailing addresses derived from the Delivery Sequence File (DSF), also referred to as the Computerized Delivery Sequence (CDS), maintained by the United State Postal Service (USPS). The frame covers all addresses receiving mail deliveries from the USPS including P.O. Boxes, addresses labeled as seasonal or vacant, and drop points (a single delivery point or receptacle that services multiple residences). It does not include non-institutional group quarters (such as college housing, group homes, shelters, and workers dormitories), unless they are coded as "residential" and have specific delivery points within the facility. Interviewing will be conducted in English and Spanish. The purpose of the address-based sample is to estimate the ownership incidence and get information about behaviors and attitudes of specific infant and toddler products with a reasonable margin of error. The address-based sample will be weighted to represent the U.S. population of households with children younger than 6 years old.

B.2 Procedures for the Collection of Information

This section describes the data collection procedures to be used in the DNPES population-based sample. The discussion includes stratification, sample selection, address-based screener and extended data collection procedures, and estimation.

Sampled residential addresses will receive a short mail screener survey with a cover letter (see also B.3.1). Based on the limited literature available (primarily an unpublished 1994 Census study by

Corteville), we plan to use an English-only mailing (letter and survey screener) for the first contact. Subsequent mailings will be English on one side and Spanish on the back. The Spanish version of the survey screener will be sufficiently different from the English version to make it difficult to use it as a guide to answering the English version. The screener survey will ask respondents for their first name, the number of children by age group who live in their household, their phone number, their email address, and their preferred follow-up mode (phone or Web). We wish to limit the number of added variables to minimize the burden on respondents, particularly given that we do not have the budget to test alternative screeners. A question about income itself will *not* be included in the mail screener, however, because it might reduce response rates. Completion of the mail screener should average 5 minutes or less. The draft of the mail screener is included as Appendix C of Supporting Statement A. The final version, if different from the draft, will be provided to OMB before the start of data collection. Each survey will also be marked with a unique identifier for tracking purposes. Returned screener surveys where the respondent indicates there are no children younger than 6 years old in the household will be finalized as ineligible.

The post-screener data collection for eligible households will be a large dual-mode investigation inquiring on infant and toddler products to understand their prevalence and usage in homes with young children. Two modes of survey administration, telephone (based on response from address-based sample) and self-administration via the Web will be used during the survey. Appendix D of Supporting Statement A contains the draft for the extended survey instrument and Appendix F contains sample screenshots of the Web survey. Minor wording changes and adjustments to skip patterns and response choices are anticipated, but the current draft contains all of the questions and content planned for the full survey and the final version will be provided to OMB when it is available and before the start of data collection. Similarly, a complete set of final Web survey screenshots will be provided to OMB before survey implementation.

The extended survey is designed to be completed in approximately 35 minutes. While administering the entire content of the questionnaire would take an excessive amount of time, each respondent will be asked about a maximum of 3 of the 24 products. The survey will execute an algorithm to choose which product "modules" will be asked of each respondent. Higher priority will be given to products that are more significant for analysis and products that are expected to have lower ownership incidence. The priorities for the selection algorithm will be updated as needed during the survey (see section B.2.5 for a more detailed description of this process). Because specific details about the features of each product are very important, the survey will include careful descriptions of each product type and, in many cases, will include the use of pictures that respondents can use for reference (these illustrations are shown in Appendix G). During preliminary cognitive testing, we found 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. For the Web survey, these reference pictures would appear within the Web survey. We want to assure that phone respondents will have access to the same information when interviewed. Therefore, for the phone survey, respondents from eligible households will receive a paper copy of these pictures by mail.

The DNPES will test a set of product usage variables that can be used to inform the CSPC about product penetration and utilization. Variables are designed to be applicable to a range of products and topics. The infant and toddler products of interest are:

- Bassinets
- Bath Seats
- Bath Tubs and Bathing Aids
- Bed Rails
- Bedside Sleepers
- Booster Chairs
- Bouncers
- Changing Tables
- Crib Bumpers
- Cribs
- Backpack Carriers with Rigid Frames
- Front Soft Carriers
- Gates
- Hand-held Carriers
- High Chairs
- Hook-on Chairs
- Play Yards
- Sleep Positioners
- Slings
- Stationary Activity Centers
- Strollers
- Swings
- Toddler Beds
- Walkers

The topics include how households acquire these products, how often they use them, the ways in which they use them, how long they use them, and what they do with these products when they discontinue using them.

Telephone surveys will be conducted by professional interviewers, using a computer-assisted telephone interview (CATI) system with a call scheduling system designed to maximize response. Prior to full fielding of the survey, interviewer training will be conducted including survey procedures and protocol, extensive training on and practice with the survey instrument, and training on gaining respondent cooperation with the survey request. Roughly 10 percent of each interviewer's work will be monitored for quality control purposes during fielding of the survey.

Web surveys will be self administered. The Web survey will be conducted using participants from the address sample who provided an email address as the preferred method of contact.

B.2.1 Respondent Universe and Sampling Frame

Recent Census data estimate that there are 110,692,000 occupied housing units in the United States (2007 American Housing Survey) and that 16,148,000 of these households (14.6 percent) include at least one child aged 0-5. Thus, the potential respondent universe is 16,148,000.

The sampling frame for the national DNPES will be constructed from a commercially available list of mailing addresses derived from the Delivery Sequence File (DSF), also referred to as the Computerized Delivery Sequence (CDS), maintained by the United State Postal Service (USPS). The advantages of an address-based sample over a random digit dialing (RDD) sampling approach are: 1) lower costs, 2) potentially higher response rates with sufficient follow up efforts to convert initial non-respondents, and 3) improved coverage (since cell-phone only households, non-telephone households, and other households not in the RDD sampling frame are included). In recent years, other Westat surveys that have used USPS-based address lists as sampling frames are the 2009 National Household Education Survey Pilot (NHES), the 2009 National Survey of Veterans Pilot (NSV), and the 2007 Health Information National Trends Survey (HINTS). In addition the National Children's Study (Montaquila et al., 2009) has used USPS address lists for quality control of traditional listing operations. In connection with an evaluation of sample design enhancements for the National Health and Nutrition Examination Surveys, Dohrmann (2007) found 97-99 percent coverage of urban areas, and 70-89 percent coverage of rural areas. Much of the low coverage in rural areas is attributable to the use of "non-city-style" addresses in these areas. However, for a mail survey, this is less of a concern since mail is expected to reach the intended recipient even though the physical location of the household may not be known at the time of sampling. Moreover, the E-911 address conversion of rural route addresses to standardized city-style addresses has improved coverage of rural areas.

The address-based sampling frame cannot be obtained directly from the USPS, but must be purchased from an authorized vendor. The frame covers all addresses receiving mail deliveries from the USPS including P.O. Boxes, addresses labeled as seasonal or vacant, and drop points (a single delivery point or receptacle that services multiple residences). It does not include non-institutional group quarters (such as college housing, group homes, shelters, and workers dormitories), unless they are coded as "residential" and have specific delivery points within the facility.

B.2.2 Sample Selection

Prior to sampling, the address frame will be sorted geographically by zip code, carrier route, and walking sequence. To obtain a representative sample of 74,074 addresses from all 50 states and the

District of Columbia, the sorted address frame will be divided into n equally-sized contiguous intervals of addresses, and one address will be selected at random from each of the n intervals.

B.2.3 Sample Size

This nationwide survey should result in completed screener interviews with approximately 16,667 U.S. households. Of these, approximately 2,500 (15 percent) should be eligible based on meeting the criterion of having a child younger than 6 years old in the household and CPSC estimates that approximately 2,000 of the eligible households will complete the full survey either by phone or Web.

With an estimated 25 percent screener response rate, the DNPES will need 66,667 valid residential addresses to yield 16,667 completed screeners. Based on experience with other recent and similar studies, approximately 10 percent of sampled addresses will be vacant, non-residential, or otherwise undeliverable. Thus, the initial sample size selected will need to be approximately 74,074 addresses to yield the target 2,000 completed extended interviews with eligible households.

B.2.4 Oversampling for Low-Income Households

The feasibility of oversampling for low-income households using 5-year American Community Survey data or Census 2010 data at the census tract or block group level, if available in time, or Census 2000 data was considered prior to an increase in project funding.

With the increased funding and the resulting increase in the 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. 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.

B.2.5 Expected Levels of Precision

If there were no variability in the weights due to weighting adjustments for screener and extended interview non-response, the expected precision in terms of 95 percent confidence interval half-widths for estimated percentages (*i.e.,* the margin of error) is given in Table B.2.5 for the two sample size options. A percentage of P=50 percent is used in the table because it yields the maximum variance of a sample-based estimate. For estimates of low-incidence products such as slings, the confidence intervals will be narrower than those indicated in the table.

Table B.2.5. 95 Percent Confidence Interval Half-Widths for a Percentage of 50 Percent

Total # Completed Interviews	95% CI Half-Width for P=50%	# Completed Interviews in Lowest Income Quartile	95% CI Half-Width for P=50% for Lowest Income Quartile
2,000	±2.2%	500	±4.5%

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 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 younger than 6 years old (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.

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).

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 section B.2.6.

B.2.6 Problems Requiring Special Sampling Procedures

The DNPES will prioritize questionnaire modules about the rarest products and the overall sample size should include enough completed surveys with each product module for meaningful analysis and reasonable precision (see section B.2.5). However, if ownership of specific infant and toddler products is significantly lower than expected, a supplementary sample of Web panel respondents purchased from a reputable vendor would provide additional information about the use of products with low ownership rates. 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. 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.

B.3 Methods to Maximize Response Rates and Address Non-Response

Based on recent similar studies conducted by Westat, the expected screener response rate is 25 percent. With an expected 80 percent completion rate from eligible screeners, the overall survey response rate is estimated to be 20 percent (25 percent x 80 percent). Each phase of the field study will employ techniques to maximize response rates and address non-response. Additionally, the DNPES will use statistical weighting to decrease bias and it will include a non-response follow-up (NRFU) effort to help estimate the level and direction of non-response bias. Please note that 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 assumed here. Early tracking of response rates will help provide better estimates, which will be used to adjust the released sample size.

Additional methods to increase response rates, such as a cash incentive or the use of FedEx for the third mailing, may be reconsidered should additional funding become available, but are not possible given the project's current funding level. 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 pre-notification, which could (at least initially) improve response rates.

B.3.1 Mail Screener

All mailed material will be simple to read and follow and as short as possible. Based on OMB comments and the available literature, we plan to use an English-only mailing (letter and survey screener) for the first contact. Subsequent mailings will be English on one side and Spanish on the back, where the Spanish version of the survey screener will be sufficiently different from the English version to make it difficult to use it as a guide to answering the English version. The screener survey will ask respondents for their first name, the number of children by age group who live in their household, their phone number, their email address, and their preferred follow-up mode (phone or Web). A final version of the screener will be provided to OMB before the start of data collection.

B.3.2 CATI Extended

To increase response to the phone survey (extended questionnaire), interviewers with experience conducting telephone surveys will be used, and multiple callbacks will be made using an automated call scheduling system. Interviewer training will focus on gaining cooperation in the first minute or so of the initial contact with a potential respondent. To maximize the contact rate, we will use a calling algorithm that handles all dimensions of call scheduling, including: time zone (respondent

and interviewer); skill level of interviewer; special needs of the case (*e.g.,* non-English language); call history; and priority of case handling.

For the telephone extended, refusal conversions will be attempted on a sub-sample of those selecting the phone as the preferred mode of communication. In addition, CATI non-respondents who also supplied an email address will be sent emails inviting them to complete the extended survey on the Web.

B.3.3 Web Extended

Web non-respondents who also provided a phone number will be followed-up by telephone interviewers and invited to complete the survey by phone.

B.3.4 Addressing Non-Response

Even surveys with very high response rates and large sample sizes can have non-ignorable non-response bias. Efforts to increase response rates to anything short of 100 percent would still leave us without an understanding of non-respondents. The plan to study non-response bias involves four strategies: (1) a non-response follow-up subsample to compare respondents who respond with additional contact effort to those who respond without the extra effort (see section B.3.4.2); (2) response propensity analysis as part of the non-response 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. Response propensity analysis and poststratification, the primary tools that will be used to understand non-respondents, are discussed in more detail below.

A non-response analysis will be done to profile screener non-respondents as part of the non-response 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 non-response 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 non-respondents. 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 non-response bias.

The non-response-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. The post-stratification factors will indicate geographic areas of under-coverage in the screener respondents (though there will be some sampling error in 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.

B.3.4.1 Statistical Weighting

Sample weights will be provided for each completed interview from the address-based sample to allow for unbiased estimation of national percentages. The sample weights are products of the base weight, non-response adjustments, and a post-stratification adjustment. The *base weight* is the reciprocal of the probability of selection of each sampled household. The *non-response adjustments* are designed to reduce the potential bias caused by differences between the responding and non-responding population and are equal to the reciprocals of weighted response rates within carefully selected response cells. The *post-stratification adjustment* modifies the non-response-adjusted base weights to the most recent Current Population Survey totals of adults by race/ethnicity, age, region of the country, and other demographic factors. This adjustment has the effect of reducing variance.

B.3.4.2 Non-Response Follow-Up (NRFU)

Data from this NRFU sub-sample will provide estimates as to whether and how response patterns differ as harder-to-contact respondents contribute the study. At the sampling stage, 9 percent of the total initial sample will be assigned to the NRFU condition. At the screener stage, sample records in the NRFU condition will receive up to 3 screener survey mailings and reminder postcards instead of just one survey mailing and one reminder. It is estimated that this effort will increase the screener response rate by 10 to 20 percentage points. Households in the NRFU condition who complete the screener and are identified as eligible will receive additional contact attempts if they do not complete the extended survey after the standard contact protocol.

Analysis of responses from the NRFU sub-sample will focus on whether and how responses from the NRFU sample change as a result of the additional contact effort when compared to the NRFU respondents who respond within the general protocol. As a random sub-sample, the NRFU data will also allow estimates (with wider variance) based on a higher response rate. These NRFU analyses will be conducted with both weighted and unweighted datasets to allow the best understanding of non-response bias and the best overall estimates.

B.3.5 Mode Effects

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 (Roberts, 2007). 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 (Dillman 2009; and Pierzchala, 2004).

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 (Dillman, 2009; Roberts, 2007; and Pierzchala, 2006).

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" (de Leeuw, 2005). 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 (Dillman, 2009).

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 (Pierzchala, 2004). Rather than presenting 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" (de Leeuw, 2005).

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" (Pierzchala, 2008). 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" (Pierzchala, 2004). 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.

B.4 Test of Procedures or Methods to be Undertaken

Once the draft survey instrument has been finalized, it will be submitted to 3 rounds of cognitive testing in both English and Spanish with up to 9 respondents in each round. Cognitive testing consists of one-on-one interviews with respondents whose key characteristics match those of the survey population (in this case, parents of children under 6 years old). After describing the purpose of the study and informing the cognitive interview respondent of her rights as a research participant, the interviewer will administer the Durable Nursery Products Exposure Survey along with a series of follow-up questions designed to understand the respondent's thought processes related to the following:

- Comprehension – do respondents understand the question as the survey designers intended?
- Recall – can the requested information be recalled and what strategies for recall are respondents using?

- Judgment and estimation processes – to what extent is the respondent motivated to take the time to accurately answer the question?
- Response processes – do the pre-coded answers to closed-end questions map accurately to the respondents' actual answers? Should the question be a closed-ended or open-ended item?

The information obtained from the follow-up questions about respondents' thought processes as they are answering the Durable Nursery Products Exposure Survey will be used to identify and refine the following:

- Instructions that are insufficient, overlooked, misinterpreted, or difficult to understand;
- Wordings that are misunderstood or understood differently by different respondents;
- Vague definitions or ambiguous instructions that may be interpreted differently;
- Items that ask for information to which the respondent does not have access; and
- Confusing response option or response formats.

In addition to cognitive interviewing, preparations will also include up to 20 total respondents over one or two rounds of usability testing in both English and Spanish for the Web survey. While completing the Web survey, respondents will be asked to comment about the appearance, interface, instructions, and layout of the Web instrument. They will be asked to identify any inconsistencies or problems with the questions displayed, as well as other issues of overall usability. Feedback from usability testers will drive refinements to the Web survey design that will help minimize respondent confusion and burden. The feedback will be used to further refine both the Web and phone surveys prior to implementation. A final version of the phone survey and complete screenshots of the Web survey will be submitted to OMB prior to implementation as well.

B.5 Individuals Consulted on Statistical Aspects and/or Analyzing Data

The individuals consulted on technical and statistical issues related to the data collection are listed below.

Jon Wivagg, Ph.D.
Senior Study Director, Westat
Phone: (240) 328-3977 Email: JonathanWivagg@westat.com

Martha Stapleton, MA.
Senior Study Director, Westat
Phone: (301) 251-4382 Email: MarthaStapleton@westat.com

Adam Chu, Ph.D.
Associate Director Statistical Area
Senior Statistician
Phone: (301) 251-4326 Email: AdamChu@westat.com

Pam Broene, MS
Senior Statistician, Westat

Phone: (301) 294-3817 Email: PamBroene@westat.com

Dr. Jill Jenkins (301-504-6795), project director, will be responsible for data collection and for analyzing the data. The individual at CPSC who will be responsible for receiving and approving contract deliverables from Westat is William Zamula.

References

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

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.

Dohrmann, S., Han, D., and Mohadjer, L. (2007). Improving Coverage of Residential Address Lists in Multistage Area Samples. Proceedings of the Survey Research Methods Section of the American Statistical Association, 3219-3226.

Montaquila, J., Hsu, V., Brick, J.M., English, N., and O'Muircheartaigh, C. (2009). A Comparative Evaluation of Traditional Listing vs. Address-Based Sampling Frames: Matching with Field Investigation of Discrepancies. *Proceedings of the Survey Research Methods Section of the American Statistical Association,* 4855-4862.

Pierzchala, M. (2006). "Disparate Modes and Their Effect on Instrument Design." In *Proceedings of the 10th International Blaise Users Conference.* Arnhem, The Netherlands: Statistics Netherlands.

Pierzchala, M. (2008). "Experiences with Multimode Surveys." In *Proceedings of Statistics Canada Symposium*. Ottawa: Statistics Canada.

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.*

Roberts, C. (2007). Mixing modes of data collection in surveys: A methodological review. National Centre for Research Methods, unpublished: http://eprints.ncrm.ac.uk/418/.

Waksberg, J. , Judkins, D., and Massey, J. (1997). Geographic-based oversampling in demographic surveys of the United States. *Survey Methodology*, 23, 61-71