Supporting Statement for:

CPSC Playground Surfaces Survey
(CPSC)

October 20, 2017

Hope Nesteruk, CPSC Contracting Office Representative
U.S. Consumer Product Safety Commission
4330 East West Highway
Bethesda, MD 20814
Telephone: 301-987-2579
hnesteruk@cpsc.gov

TABLE OF CONTENTS

B. Collection of Information Employing Statistical Methods

B.1. Potential Respondent Universe and Sampling Methods.

Potential Respondent Universe

The target population is U.S. households with children ages 0–5 years. According to recent data from the U.S. Census Bureau, there are an estimated 115,852,000 occupied housing units in the United States (2013 American Housing Survey), of which 16,238,000 have at least one child ages 0–5 years. Thus, the potential respondent universe is 16,238,000 households. Within sampled households, the set of potential respondents is English- or Spanish-speaking adult parents or guardians of children ages 0–5 years.

Sampling Methods

Overview

With the challenges of the low incidence population in mind, the CPSC sample plan will capitalize on the ability to re-contact potential respondents who had been previously reached via a dual-frame RDD sampling design, namely, the SSRS Omnibus survey. This sample source will increase the cost-efficiency of data collection and reduce survey burdens, given that the use of previously collected information (from the SSRS Omnibus) to identify potential respondents for the current survey effort will lead to higher contact rates and higher screening eligibility rates.

Note that while the sampling design aims to minimize survey error to the extent possible within cost and other practical constraints, this study is not designed with the intent of generating nationally representative data. Achieving a nationally representative sample with a high response rate and equivalent sample size to that of the proposed design would be much more expensive, and is unnecessary for purposes of achieving the goals of this study.

SSRS Omnibus

The SSRS Omnibus is a national, weekly, dual-frame bilingual telephone survey. Each weekly wave of the SSRS Omnibus consists of 1,000 interviews, of which 600 are obtained with respondents on their cell phones, and approximately 35 interviews completed in Spanish.

The SSRS Omnibus entails interviews with adults in the U.S. (including Hawaii and Alaska). SSRS Omnibus uses a fully-replicated, single-stage, random-digit-dialing (RDD) sample of landline telephone households, and randomly generated cell phone numbers. Sample telephone numbers are computer generated and loaded into on-line sample files accessed directly by the computer-assisted telephone interviewing (CATI) system. The SSRS Omnibus uses an overlapping dual-frame design, with respondents reached by landlines and cell phones. The RDD landline sample was generated through Marketing Systems Group's (MSG) GENESYS sampling system. The standard GENESYS RDD methodology produces a strict single stage, Equal Probability Selection Method (epsem) sample of residential telephone numbers. In other words, a GENESYS RDD sample assures an equal and known probability of selection for every residential telephone number in the sample frame, prior to nonresponse. The sample is generated shortly before the beginning of data collection to provide the most up-to-date sample possible, maximizing the number of valid telephone extensions. Following generation, the RDD sample is prepared using MSG's proprietary GENESYS IDplus procedure, which identifies and eliminates a large percentage of all non-working and business numbers.

Using a procedure similar to that used for the landline sample, MSG generates a list of cell phone telephone numbers in a random fashion. Inactive numbers are flagged and removed using MSG's CellWins procedure.

Within each landline household, a single respondent is selected through the following selection process: First, interviewers ask to speak with the youngest adult male/female at home. The term "male" appears first for a random half of the cases and "female" for the other randomly selected half. If the requested adult male/female is not available (*e.g.*, in a single parent household, or parent is not at home), interviewers ask to speak with the youngest adult female/male (asking for the other gender) at home. The SSRS Omnibus asks for the youngest adult in the household in order to yield more interviews with younger adults, who tend to have higher rates of nonresponse. This method of within-household selection of sample members should improve the balance of the responding sample with respect to age, which can be expected to reduce weight variability and thereby improve precision.

Cell phones are treated as individual devices and the interview may take place outside the respondent's home; therefore, cell phone interviews are conducted with the person answering the phone.

During the SSRS Omnibus interview, detailed demographic data is collected for each respondent, including age, gender, marital status, and number and age of child in their household.

<u>Study-Specific Sampling Procedures</u>

We will pull the target sample for the CPSC survey from the pool of respondents to the SSRS Omnibus. We plan to use data from the past three years of the Omnibus Survey (2015 through 2017). We will select two types of Omnibus respondents to re-contact for this study:

- *Households with Children***:** Respondents in households with a child age 0 to 5;
- *Potential New Families***:** Respondents in households without children, but with an adult age 18 to 34[1] who is married or cohabitating. We have included this sample source to improve the number of first-time parents with newborns in the sample, as the Omnibus sample will be "aging" and therefore less likely to include these new families.

We will re-contact telephone households who meet these criteria and screen them for qualification into the study (parents/guardians of child 0–5). We assume that all

[1] The current proposal will use data from the previous three years. A 34-year old mother who took the survey would be 37 now, and most women (90%) have their first child by age 37; https://www.cdc.gov/nchs/data/databriefs/db232.pdf.

parents/guardians of at least one child age 0–5 will qualify for the study (100% incidence rate among this group).

Two forms of within-household random selection will be applied. For selecting a sample member to be interviewed, we will randomize whether the interviewer asks for an *adult male* or *adult female* parent or guardian of a child 0–5; if the adult of the specified sex is not available (*e.g.*, in a single parent household), then we will ask whether an adult of the other sex is available. The term "male" appears first for a random half of the cases and "female" for the other randomly selected half. For eligible sample members with more than one child in this age range, the child with the most recent birthday will be selected to be the referenced child for specific questions in the survey. Note that these methods may be considered pseudo-random rather than truly random, but will reduce respondent burden and cost and increase the response rate compared with a rostering method, while also achieving a reasonable degree of randomization.

With respect to within-household selection, it should be noted that although we are sampling adult parents and guardians of a child 0–5 (*i.e.*, the survey population), we are randomly selecting members of the survey population within household, rather than specifically asking to speak with an individual who accompanies his or her child to the playground. Although this subpopulation (*i.e*., parents or guardians of a child 0–5 who accompany their child to the playground) is of high interest, there is no source of external benchmarks for this subpopulation, which would prevent the ability to compute valid weights that correspond with a well-defined target population. Further, the within-household substitution of such subpopulation members in place of members of the survey population who are outside of this subpopulation could be expected to introduce a selection bias that would be difficult to quantify and/or to mitigate. Our proposed method avoids these pitfalls.

Anticipated Response Rates
Among Omnibus respondents invited to participate in the CPSC study, we anticipate a study-specific completion rate of roughly 30%. This estimate is based on similar studies that also entailed sampling from SSRS Omnibus respondents. Note that this completion rate is a measure of study efficiency and of nonresponse in the final stage (among Omnibus respondents who were invited to participate in the CPSC's studies), but does not reflect Omnibus-specific nonresponse, and is therefore not a population-level response rate. Typically, Omnibus attains a response rate of 5%–7% (American Association for Public Opinion Research Response Rate 3 [AAPOR RR3]; AAPOR, 2016).[2] Thus, we anticipate a population-level response rate, reflecting both phases of nonresponse (*i.e.,* Omnibus nonresponse and study-specific nonresponse), of roughly 2%–3% (AAPOR RR3; computed as the product of the two previously mentioned rates: the Omnibus response rate and the study-specific completion rate).

B.2. Procedures

Statistical Methodology for Stratification and Sample Selection
The statistical methodology for sample selection is described above in section B.1.b. In order to achieve adequate precision while also using the most recently available Omnibus data as a

[2] The American Association for Public Opinion Research. 2016. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition. AAPOR.

starting point for the CPSC study, we expect to use a sampling rate of 100% of Omnibus respondents. Therefore, there would not be any additional subsampling of households, nor stratification of Omnibus respondents, but rather, we would re-contact all Omnibus respondents who qualify for sampling for this study (*i.e.*, households with children, and potential new families, as described above, and who responded during the specified time period). Subsequently, within-household sample selection would be applied, as described in section B.1.b above, in order to randomly select the respondent (*i.e*., adult male or adult female parent or guardian of a child 0–5 years old) and randomly select the child to be referenced for specific questions in the survey.

Estimation Procedures

Estimates will be produced using standard survey estimation procedures for complex sample designs. These procedures are based on a design-based, model-assisted paradigm for statistical inference (e.g., Särndal et al. 1992; Valliant et al. 2013).[3,4] Estimation procedures will include the use of survey weights to account for the study design and mitigate the risk of various sources of survey error. Survey estimates of interest include estimates of child behaviors that expose them to playground surfacing. Exposure may include skin contact, ingestion, and potential contact through open wounds. A particular sub-group of interest includes parents who take their children to playgrounds with recycled tires for filling. Variance estimates will be computed via Taylor series linearization and will be computed via an appropriate software package (*e.g.*, SAS, SUDAAN, WesVar, or Stata).

The survey weights will account for the fact that not all sample members were selected with the same probabilities and will be adjusted to account for systematic nonresponse along known population parameters. We expect the weighting to involve several stages:

1. <u>Adjustment for likelihood of selection and successful re-contact efforts (base-weight)</u>**.** This will be based on:
    a. Probability of phone selection: A phone number's probability of selection depends on the number of phone-numbers selected out of the total sample frame. For each respondent whose household has a landline phone number, this is calculated as total landline numbers dialed divided by total numbers in the landline frame and conversely for respondents answering at least one cell phone number, this is calculated as total cell phone numbers divided by total numbers in the cell phone frame.
    b. Probability of contact: For the landline sampling frame, the probability that a given household is selected is proportional to the number of eligible landlines in that household (e.g., a household with two working landlines has twice the probability of selection than a household with one working landline). For the cell phone sampling frame, the probability that a given individual is sampled is proportional to the number of eligible cell phones owned by that individual.
    c. Probability of selection within household (landline frame only): In households reached by landline, only one adult is selected. If selection were completely random

[3] Särndal, C. E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*.
[4] Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York: Springer.

within household, then the probability of selecting a given adult within the household would be inversely related to the number of adults in the household, after accounting for other factors (e.g., number of landlines).[5] Thus, it is necessary to account for the number of adults within households to avoid underrepresentation of adults who live in multi-adult households. Use of multiple sampling frames: The use of separate sampling frames for landlines and cell phones necessitates accounting for the combining of two frames, and to reflect that some households can be selected from either frame (*i.e.*, individuals who can be reached via both landline and cellphone), which affects the probability of selection. This will make sure that individuals who can be reached via both sampling frames are not overrepresented.

d. Propensity to respond: Systematic non-response based on the fact that some pre-screened Omnibus respondents will be successfully re-contacted and others will not. Propensity weights will rebalance successfully re-contacted pre-screened respondents to the original sample pool of pre-screened Omnibus respondents. The propensity weight will use a standard logistic regression procedure, leveraging the over 25 demographic benchmarks attained in the omnibus survey. A backwards entry procedure will reduce the model to variables at least minimally significant to the propensity for a successful re-contact. If deemed necessary to reduce variance, predicted propensities will be reduced to a five-level weighting class variable.

More specifically, we recommend accounting for aspects 1(a)—1(d) above in a single step via methods outlined by Buskirk and Best (2012; equation 3).[6] This procedure is motivated by the addition rule of probability, which follows from the inclusion-exclusion principle.[7] This method computes base weights as follows:

$$\mathrm{bw}_{\{\mathrm{SF-BP\,(OLC)}\}} = \big(\mathrm{P}(\mathrm{in\ S})\big)^{-1} = \left(\left(\frac{\mathrm{S_{LL}}}{\mathrm{U_{LL}}} \times \frac{\mathrm{LL}}{\mathrm{AD}}\right) + \left(\frac{\mathrm{S_{CP}}}{\mathrm{U_{CP}}} \times \mathrm{CP}\right) - \left[\left(\frac{\mathrm{S_{LL}}}{\mathrm{U_{LL}}} \times \left(\frac{\mathrm{LL}}{\mathrm{AD}}\right) \times \frac{\mathrm{S_{CP}}}{\mathrm{U_{CP}}} \times \mathrm{CP}\right)\right]\right)^{-1}$$

where the terms are defined, as follows:

Landline Frame	Cell Frame
$\mathbf{U_{LL}}$= the size of the landline sample frame	$\mathbf{U_{CP}}$ = the size of the cell sample frame
$\mathbf{S_{LL}}$ = the amount of landline sample released	$\mathbf{S_{CP}}$= the amount of cell sampled released
LL= the number of landline telephones in the household that are used to receive calls	**CP**= The number of cell phones owned by the respondent (or more simply whether or not the adult has a cell phone).
AD= the number of adults in the household (assumed to be at least 1)	

[5] As previously noted, the Omnibus uses a quasi-random method of selection within household (i.e., asking for the youngest adult male/female), in order to improve representation of younger adults. Therefore, younger adults have a higher probability of selection within household than older adults, conditional on other factors (e.g., contact). However, younger adults tend to respond to surveys at lower rates, and in telephone surveys, person-level base weights can only be computed for respondents. Therefore, in computing the base weights, an implicit assumption is made that the higher probability of selection within household for contacted younger adults is canceled out by their lower likelihood to be contacted. Possible bias due to violation of this assumption may be mitigated via the subsequent calibration adjustments.

[6] Buskirk, T. D., & Best, J. (2012). "Venn Diagrams, Probability 101 and Sampling Weights Computed for Dual Frame Telephone RDD Designs." In *Proceedings of the American Statistical Association*, Survey Research Methods Section.

[7] More specifically, the inclusion-exclusion principle, as applies to events A and B in a probability space, indicates that $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.

Buskirk & Best further recommend topcoding CP at 3, LL at 2, and AD at 4, which will reduce weight variation and may simplify the required question wording; they indicate that these caps typically affect at most 4%—5% of the sample. Their procedure has conceptual similarities to a single-frame approach to dual-frame estimation, and does not require a compositing adjustment.

Note that the base weights via this method can only be computed for the set of Omnibus respondents, rather than the full set of invited sample members, given that several of its elements are unavailable for nonrespondents. Thus, nonresponse to the Omnibus survey is implicitly handled via the subsequent calibration step.[8]

After computing the base weights via Buskirk & Best, step 1(e) above will be applied to account for nonresponse to the CPSC study among survey invitees (*i.e.*, Omnibus respondents), either by a response propensity adjustment or via response propensity stratification. If a response propensity adjustment is applied, the weight of a given responding sample member will be multiplied by the inverse of the sample member's model-estimated probability of response, whereas the weights of nonrespondents will be removed. If response propensity stratification is used, then the estimated response propensities will be used to form five weighting classes; within each weighting class, an adjustment factor will be computed as the total weights of all sample members (in class) divided by the total weights of respondents (in class), whereas the weights of nonrespondents will be removed.

2. Calibration weighting (raking)**:** With the base-weight applied, the sample will be balanced to reflect the distribution of the adult population who were the parents/guardians of at least one child aged 0 to 5 along known population parameters.

   The balancing will be done using iterative proportional fitting (or 'raking'), a procedure in which the weights are repeatedly adjusted to the control totals until the difference between the weighted data and the population benchmarks is near zero.

   The demographic benchmarks will be based on the most recent available U.S. Census Bureau's American Community Survey (ACS). The parameters likely to be used are:

   a. age of parent (18–24; 25–29; 30–49; 50–64; 65+),
   b. gender (male; female),
   c. education (high school or less, some college, four-year college, graduate degree or more),
   d. race/ethnicity (White non-Hispanic; Black non-Hispanic; Hispanic; Other non-Hispanic),
   e. marital status (married; not married),

[8] Alternatively, a multi-step weighting procedure could be designed that would explicitly account for Omnibus nonresponse, while also accounting for other design aspects. However, given the lack of auxiliary variables in the original RDD sampling frames and typical similarity of response rates for landline and cell phone frames in the Omnibus surveys, we do not foresee any meaningful benefits from such an approach, while the added complexity would likely increase weight variation and reduce precision. Nevertheless, if eligibility and/or response rates are meaningfully different by sampling frame, we will assess whether there may be benefits to modifying these procedures and/or incorporating telephone usage in the calibration benchmarks.

f. region (Northeast; Northcentral; South; West), and
g. age of child (0–2; 3–5).

3. Trimming**:** Adjustments are then made to control the variance of weights ('trimming'), constraining weights typically to top/bottom ≤ 5%, depending on the specific outcomes.

As a whole, the set of weighting procedures will result in a single set of weights for survey respondents. The weights aim to mitigate various sources of survey error to the extent possible given the limitations of the sample, while allowing for sample-based estimates that conform to external benchmarks for a well-defined target population. However, it should be noted that the study, which entails sizable levels of nonresponse, is not designed to create nationally representative estimates.

Degree of Accuracy
After accounting for the study design, we anticipate that the set of 2,200 completed interviews will result in a margin of error (MOE) of approximately 2.8%, with a 95% confidence level, assuming a response proportion of 0.5 and a design effect from weighting of about 1.8. The design effect from weighting is computed as 1 plus the squared coefficient of variation of the weights, and reasonably approximates the design effect (DEFF) for single-stage designs when the weights are not correlated with the survey variable being estimated (Spencer, 2000).[9] The anticipated design effect from weighting of 1.8 is based on SSRS's experience with similar studies from among Omnibus respondents, which typically have design effect from weighting of 1.7–1.8, and occasionally 1.9. Note that the estimated margin of error above is for the full set of interviews; in practice, precision will be further reduced for subpopulation estimates. Standard error estimates will reflect the survey weights and complex sample design and will be computed using Taylor series linearization.

It should be noted that results will not be used to infer highly precise point estimates, but rather, to obtain descriptive information about the target audience and to inform regulations regarding the use of potentially toxic material. We believe that the sample size and design will allow for sufficient precision for key quantities being estimated, including for subpopulations, while also providing anticipated benefits commensurate with the survey costs and anticipated respondent burden.

It should also be noted that although Taylor series linearization is a well-accepted method for variance estimation in many sample survey contexts, this study is not designed to generate nationally representative data. Therefore, such variance estimates and any associated measures of precision (e.g., MOE) will not reflect all sources of non-sampling error, such as coverage bias and/or nonresponse bias. However, obtaining highly accurate measures of precision are not necessary to achieve the goals of the project.

B.3. Maximizing Response Rates.

Maximizing Response Rates

---

[9] Spencer, B. D. (2000). "An approximate design effect for unequal weighting when measurements may correlate with selection probabilities." *Survey Methodology*, *26*(2), 137–138.

Re-contacting participants who have completed the Omnibus survey and fit the target criteria (parents of children who are currently 0–5 years old) will result in a high participation rate and efficient data collection, although the cumulative response rates will be low due to earlier phases of nonresponse. In an effort to maximize the response rates, respondents are given every opportunity to complete the interview at their convenience. For instance, those refusing to continue at the initiation of or during the course of the interview will be offered the opportunity to be re-contacted at a more convenient time to complete the interview. Non-responsive numbers, such as no answers, answering machines and busy signals, receive six call attempts.

A key way to increase responses rates is through the use of refusal conversions. Phone interviewers will be highly experienced in refusal conversion, and will redial all initial refusals on this project to attempt to convert them to final completed interviews.

Implications of Nonresponse, as Relating to Survey Weights

As per Little & Rubin (2002), the modern statistical literature distinguishes between three types of missing data: data that are missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR).[10] Methods for accounting for unit-nonresponse in surveys via weighting, both in this survey and more generally, typically assume that the mechanism for unit-missing data is MAR—that is, that conditional on observed characteristics, that the data missingness is independent of the outcome measures; this is a weaker assumption than MCAR. This assumption is often made implicitly, and can be used to motivate the use of response propensity adjustments for nonresponse (*e.g.*, explicit model, as in weighting step 1e above) and the use of calibration adjustments *(e.g.*, implicit model implied by weighting step 2 above). Assuming that models used in weighting (whether implicit or explicit) take advantage of key auxiliary variables and appropriately reflect the patterns of missing data, then such adjustments can be effective at mitigating selection bias.

Unfortunately, it is typically difficult or impossible to assess whether unit-missing data are NMAR (*e.g.*, Valliant et al. 2013, p. 319).[11] If the data are NMAR that indicates that the data missingness is not independent from unobservable characteristics, even after accounting for the observable characteristics. However, such unobservable characteristics are, by their nature, not observed for nonrespondents.

For this survey, we will assume that the unit-missing data are MAR. This is primarily out of necessity, as explained above, given the inability to adjust based on unobservable characteristics without making potentially strong assumptions. This assumption is typically made when computing survey weights. However, we also note that use of several sociodemographic characteristics for weighting adjustment purposes should mitigate the risk of error, and in conjunction with the planned data collection methods, should yield estimates of adequate quality, particularly given that this study is not designed to be nationally representative, nor does it need to be.

[10] Little, R. J., & Rubin, D. (2002). Statistical analysis with missing data.
[11] Valliant, R., Dever, J. A., & Kreuter, F. (2013). Practical tools for designing and weighting survey samples. New York: Springer.

If evidence arises as to suggest that data missingness patterns are NMAR, then we may conduct sensitivity analyses to assess the possible impact of different types of models, or we may conduct intensive modeling efforts for outcomes of particular interest (*e.g.*, via sequential regression through multiple imputation, which may handle more complex data missingness patterns). However, we think it is unlikely that such intensive modeling efforts will be necessary for this particular survey, given that the proposed data collection methods are expected to produce estimates of sufficient fitness for the purposes for which they are being used.

Nonresponse Bias Analysis

For this study, we will conduct two types of nonresponse bias analyses during the course of computing survey weights, which aim to mitigate possible nonresponse bias.

First, we will conduct an auxiliary variable analysis as part of computing nonresponse weighting adjustments in weighting step 1(e). This analysis will focus on the last phase of nonresponse (*i.e*., nonresponse to the CPSC study among Omnibus sample members), which allows for the use of Omnibus survey responses as predictors for subsequent nonresponse. Logistic regression methods will be used to estimate sample members' probability of response to the CPSC study, among Omnibus respondents, using variables obtained during the Omnibus data collection. A statistically significant model would suggest that the unit-missing data may not be MCAR and may lead to nonresponse bias in unadjusted estimates for survey variables that are correlated with any statistically significant predictors. Therefore, this would help motivate the previously described response propensity adjustment (or variant thereof).

Second, in the course of computing calibration weighting adjustments for weighting step (2) above, we will conduct benchmarking analyses to assess differences between sample-based estimates and external benchmarks. However, note that weight calibration ensures conformity between the weighted sample and external benchmarks with respect to the weighting adjustment categories. Therefore, these benchmarking analyses will primarily be applicable during the course of designing the calibration weighting dimensions, rather than in assessing bias of the calibrated estimators. For example, a benchmarking analysis that exhibits meaningful differences between weighted estimates and benchmarks (*e.g.*, for a variable not used in adjustment or larger set of categories than were used in adjustment) may suggest possible benefits to modifying the adjustment categories (subject to bias-variance tradeoffs that may result from increased weight variation). These analyses may also inform decisions related to weight trimming (*e.g.*, whether to re-calibrate the trimmed weights).

B.4. Pretesting

FMG and SSRS will test the survey instrument to identify comprehension issues, programming errors, inaccurate skip patterns, and internal logic issues. Any substantive changes following public comments gathered during the OMB public review period will be submitted to OMB.

B.5. Data Collection

The survey will be conducted by SSRS.

Jordon Peugh
(862) 252-4235