



**UNITED STATES
CONSUMER PRODUCT SAFETY COMMISSION**

This document has been electronically
approved and signed.

Memorandum

Date: May 20, 2020

TO: The Commission
Alberta E. Mills, Secretary

THROUGH: Mary T. Boyle, Executive Director
DeWane Ray, Deputy Executive Director for Safety Operations
Duane Boniface, Assistant Executive Director
Office of Hazard Identification and Reduction
Joel Recht, Deputy Assistant Executive Director
Office of Hazard Identification and Reduction

FROM: Steve Hanway, Associate Executive Director
Directorate for Epidemiology

SUBJECT: Additional Data Sources for CPSC

Background

CPSC's Fiscal Year 2020 Operating Plan directs the Office of Hazard Identification and Reduction (EXHR) to "identify, research, and evaluate other sources of data and data intake systems." This memorandum summarizes the data sources and systems that were evaluated.

We identified possible new data sources by first identifying gaps in existing CPSC data. We then identified possible data sources to address these gaps, researching the potential sources, and evaluating their potential utility and value to CPSC, relative to potential costs.

Existing CPSC Data

Existing CPSC data sources are useful and informative for meeting mission demands, with data covering a range of incident types, as well as consumer injuries and fatalities. CPSC major sources of data include the National Electronic Injury Surveillance System (NEISS), incident reports found in the Consumer Product Safety Risk Management System (CPSRMS), and data that CPSC acquires from the National Fire Incident Reporting System

(NFIRS) and the Centers for Disease Control (CDC). CPSC staff uses these data to identify hazards to consumers and evaluate the risk to them from products, as well as to evaluate the effectiveness of interventions to improve safety. The following is a brief overview of some of the primary sources.

NEISS

NEISS is a probabilistic survey of Emergency Departments (EDs). NEISS data allow CPSC to make national estimates of injuries associated with consumer products, identify changes in injury rates to help detect emerging hazards, and evaluate the effectiveness of rules and voluntary standards. Also, staff uses the data to determine the societal costs of consumer product injuries (using CPSC's Injury Cost Model). In some cases, staff uses NEISS to contact injury victims to conduct special studies or collect additional incident information.

NEISS data is limited, however, by several factors. The details provided are generally restricted to information that medical personnel deemed noteworthy for their medical records. Also, NEISS cases typically do not include manufacturer information. Furthermore, NEISS data do not include most fatalities, because only a minority of fatalities involve ED visits.

CPSRMS

CPSRMS includes anecdotal reports from numerous sources including death certificates, medical examiners and coroners, manufacturers and retailers, newsclips, consumers (*e.g.*, SaferProducts.gov), referrals from other government agencies, healthcare providers, public safety entities, and CPSC's in-depth investigations.

CPSRMS data contribute to CPSC's understanding of potential consumer product hazards. Collectively, CPSRMS data provide in-depth information not typically found in the NEISS, such as manufacturer and product information, in-depth incident details and background, photographic evidence and, in many cases, contact information to enable outreach to victims for investigation. The detailed description of hazard scenarios can often make clear the failure mode associated with the product in question. Even without the assurance of inclusiveness of every incident with a product, CPSRMS data provide a count of minimum known incidents which is often a sufficient basis for agency action.

There are limitations to CPSRMS data, as individual incidents often lack details. In addition, because the set of data is not statistically representative, it cannot be used to develop national estimates or annual trends.

NFIRS

The National Fire Incident Reporting System, administered by the U.S. Fire Administration, collects information in a uniform manner from fire departments about the fires to which they respond. When combined with information from the National Fire Protection Association's national fire loss estimates and adjusted to account for errant

outliers using CPSC's in-depth investigations of reports of multiple fatalities and high property loss, this data can be used to produce estimates of fires associated with numerous products under CPSC's jurisdiction.

CDC

The Centers for Disease Control and Prevention (CDC) makes the Public Use Mortality Multiple Cause File from the National Vital Statistics System available annually via download. The file provides summary information on all deaths in the United States. This information is useful to CPSC in cases where knowing the International Classification of Diseases, 10th Revision (ICD-10) is useful. This also provides a basis for CPSC to extrapolate from the set of death certificates the Commission receives to every known death certificate known to share an ICD-10 code for a particular year. This allows CPSC to make estimates from its own death certificates of consumer product-associated deaths.

Gaps in Existing CPSC Data

There are circumstances where additional information would be helpful to provide sufficient understanding of a specific risk, hazard, or product. The current data can answer numerous questions. Nevertheless, there are questions that these data cannot answer.

Among the gaps identified in CPSC's current data were:

- Social media/online reviews
- Medical treatment outside of emergency departments
- Death reporting
- Exposure/sales

Social Media/Online Reviews

Communication via social media has become an increasingly common approach to sharing information. Credible news sources have been important contributors to CPSC's monitoring of consumer product incidents that may result in injury or fatality. However, today's media environment is one in which the prominence of established news outlets is diminished, with public attention shifting more and more to alternatives. Additionally, user comments on websites that sell consumer products may contain useful information concerning negative experiences with specific products, including near-miss information that may not be reported as part of Section 15. Monitoring these sources in a systematic way may allow CPSC to identify emerging hazards earlier.

Medical Treatment Outside Emergency Departments

The National Electronic Injury Surveillance System (NEISS) collects data on visits to Emergency Departments (EDs). The rise in the use of urgent care centers (UCCs) has raised questions in the past about the utility of seeking injury data for injuries treated in non-hospital environments though a recent evaluation of this option found little current utility. Other settings include inpatient environments, long-term care environments, doctor's offices, health centers and ambulatory surgery settings.

Death Reporting

CPSC does not receive a copy of every death certificate for each year. Although CPSC augments its understanding of consumer-product fatalities with Medical Examiners and Coroners Alert Project (MECAPS) reports, newsclips, in-depth investigations, and reports from consumers and manufacturers, it is not assured it will receive a report of every consumer product-associated fatality that occurs in the United States.

Exposure/Sales

CPSC's current data do not typically answer fundamental questions about product ownership, use, and exposure. When assessing the risk associated with a consumer product, it is meaningful to consider not just the known instances where an injury or fatality has occurred, but also the instances in which a product was used or available to the consumer and no injury occurred. Although staff can often obtain information on the number of products sold, and can estimate the number of products available for use (given product life estimates), dimensions, such as frequency and duration of use, are not available. These dimensions have significant impact on risk, and without them, staff is forced to make significant assumptions about product use. Exposure information may also play an important role in understanding chronic hazards, where frequency and duration of use may need to be considered as a function of time and different stages of life, to estimate impacts from consumer products on end-points like cancer.

Potential New Sources of Data

CPSC staff explored a number of possible data sources, including:

- Point-of-sale data
- Web-crawling or Web-scraping data
- Urgent care center data
- HCUP/AHRQ data

- CDC data
- Child Death Review Board data
- Poison Control Center data.
- Exposure surveys

Point-of-Sale Data

CPSC staff is aware of point-of-sale data that could be used to understand the products that consumers purchase. These data sets are mostly obtained by compiling scanned bar code information at point of sale checkouts. CPSC staff could use point-of-sale data to contextualize our existing injury or fatality information in terms of products sold. In this way, point-of-sale information would enable a comparison of the number of incidents to the number of products sold, which would provide a measure of the risks associated with particular products.

Changes in the number of products sold can be an important component of change in the number of injuries observed. For example, staff observing a constant number of injuries while unaware of growing sales, might not recognize effectiveness of worthwhile safety measures. Likewise, if staff is not aware of a declining sales environment, and only sees a constant number of injuries, this might mask an increase in the rate of injuries occurring with a particular product.

Benefits

The data would allow CPSC to understand the actual sales of consumer products, based on direct observation at the point of sale. Furthermore, staff could make use of specific information on the manufacturer and model of product involved to aid in investigations. Although this data does not provide as rich information on usage as a household exposure survey would, the data sets exist and are commercially available from at least one major vendor.

Concerns

CPSC may have difficulty in obtaining this data due to unwillingness of vendors to provide it to us. One issue is that vendors of this data rely on sales of the data for income. Thus, if CPSC were required to release these data in response to FOIA requests, the public may gain free access to the vendor's data. A mechanism for protecting the information would likely be necessary to find a willing provider.

Furthermore, there may be some disincentive for point-of-sale vendors to provide sales data to CPSC, if the manufacturing firms that make up their traditional customer base considered the data potentially detrimental to their interests. The major providers in this area have not been responsive to staff inquiries to date, perhaps because of the two concerns discussed

above. Given the lack of vendor responsiveness, staff cannot estimate the potential costs to acquire the data.

Recommendation: Authorize further study of this option, including legal analysis by OGC, market research, and a potential request for a price quote (RFQ).

Web-crawling or web-scraping data

Certain tools exist, or could be constructed, to harvest information from social media and/or online retail that may be of interest to the Commission. For example, tools exist for determining the average cost of a product, by visiting sites that sell it.

Keyword searches and sentiment analysis are possible for published sources of information like Twitter. CPSC could use these tools to create algorithms to retrieve posts that mention certain keywords. This might permit CPSC to gain aware of incidents with consumer products that would not appear within our existing data sources. We would likely need a highly capable automated system for such an approach along with artificial intelligence/machine learning tools to extract the useful information.

Benefits

The major benefit would be to provide CPSC a monitoring capability of social media. There may be instances where this is the only means where CPSC could become aware of a particular consumer product safety injury or incident or at least become aware more quickly.

Concerns

In some cases, the tools require agreement with terms of service for these websites, which may have language restricting the allowable uses of the information displayed.

Given the volume of information on social media, the variety of ways language can be used, and the limited specificity of the information provided, it is likely that most of the content retrieved would not be useful for determining actions or interventions for the CPSC or CPSC staff. This is because the most useful information would have specificity about the product and hazard scenario and would allow CPSC to make contact with the victim or their next of kin to conduct an in-depth investigation. Web-crawling data tends to lack some or all of these attributes.

If large volumes of this information were stored locally by CPSC, it might entail additional organizational costs for storage and security. A high volume of information might also require a great deal of staff time to understand and monitor.

A recent inquiry into the costs for obtaining web-scraping information from a third party service for a single class of products found commonly in homes was approximately \$250K. If used frequently, the costs could be substantial.

Recommendation: Do not pursue this at the current time. Instead, continue market research and monitoring of evolving tools and practices and continue development of machine learning tools for dealing with high volumes of unstructured data.

Urgent Care Center Data

An informational memorandum¹ went to the Commission on April 8th, 2020 from the Directorate for Epidemiology. The “Supplemental Analysis of the Value of Data from Urgent Care Centers for CPSC” considered and discussed this option. The memorandum did not find a large-scale collection of data from UCCs to be advisable, as UCC’s were not an important treatment setting for severe injuries at this time.

Recommendation: Collection of data from UCCs does not seem advisable at the current time. CPSC staff should continue to monitor publicly information about injuries seen in UCCs. This is described more fully in Staff’s UCC report.

HCUP/AHRQ Data

The Agency for Healthcare Research and Quality (AHRQ) oversees the Healthcare Cost and Utilization Project (HCUP). HCUP databases bring together the data-collection efforts of state data organizations, hospital associations, private data organizations, and the federal government, to create a national information resource of encounter-level healthcare data. These databases provide national- and state-level information on care, delivered in a variety of settings. Some of these (MEPS, NEDS, NIS, SID) are included as inputs to CPSC’s Injury Cost Model. Numerous databases are made available for purchase or download, including:

National Inpatient Sample (NIS) - Inpatient utilization, charges, and outcomes

Kids Inpatient Database (KID) – Robust sample of pediatric inpatient care

Nationwide Ambulatory Surgery Sample (NASS) – Major Ambulatory Surgeries

Nationwide Emergency Department Sample (NEDS) – All cause ED visits/charges

Nationwide Readmissions Database (NRD) – Informs estimates of readmission rates

State Inpatient Databases (SID) – Inclusive inpatient data from community hospitals

State Ambulatory Surgery and Services Databases (SASD) – Rich surgery data

¹ https://www.cpsc.gov/s3fs-public/Supplemental%20Analysis%20of%20UCC%20Data.pdf?GRbUy0NpQHDZh9icxjTw_X4Efa4FAQRs

State Emergency Department Databases (SEDD) – ED visits without hospitalizations

Medical Expenditure Panel Survey (MEPS) – Cost and use of healthcare.

Benefits

The HCUP and AHRQ data can be a useful to quantify specific types of healthcare treatment and costs. When CPSC is considering treatment costs as a component of societal costs, this could allow information on specific treatments to be considered. Most of the data sets are affordable (costing \$1,000 per year or less). Further, the geographic specificity could be a useful input for small area estimation, a technique for making estimates about smaller geographies from larger geographic data sets.

Concerns

The data are not specific to consumer products, so despite their specificity to the costs of treatments, there may be difficulties identifying the cases most relevant to CPSC. Despite the fairly modest cost of the data, the staff time needed to overcome these challenges relating to specificity of cause make use of the data and learning curve in gaining familiarity with them may not be trivial.

Recommendation: Procure and employ these data, when needed to answer specific questions of importance to CPSC that cannot be answered with CPSC data alone. Given the low costs, these data can be purchased without large budget allocations in advance.

CDC Data

The Centers for Disease Control and Prevention (CDC) includes the National Center for Health Statistics (NCHS), which collects a variety of medical treatment data. Some of these data are restricted and require submission of a research proposal to gain access. The billing cost for the restricted data is determined based on the level of effort required to prepare the data for use. Their mortality data would be particularly useful if the CDC agrees (via a MOU) to provide access to the narrative information. Numerous databases are made available for purchase or download, including:

National Ambulatory Medical Care Survey (NAMCS) – Doctors’ office/health center visits

National Hospital Ambulatory Medical Care Survey (NHAMCS) – Utilization of EDs, outpatient, and ambulatory surgery locations [Not collected 2018-2020]

National Health Interview Survey (NHIS) – Household survey of health in United States

National Hospital Care Survey (NHCS) [Restricted] – Trends affecting hospitals
National Study of Long-Term Care Providers (NSLTCP) [Restricted] – Monitors trends in paid, regulated, long-term care services

Benefits

These data can be a useful quantifier of specific types of healthcare treatment and costs which might be useful for computing societal costs of certain kinds of consumer product injuries. Some of these data can be obtained at no cost. The restricted use data cost is expected to be comparable to the data provided by HCUP or AHRQ, although these costs are determined after access is granted and thus not able to be determined in advance of that.

Concerns

The data are not specific to consumer products so despite their specificity to the costs of treatments, there may be difficulties identifying the cases most relevant to CPSC. Despite the fairly modest cost of the data, the staff time needed to overcome these challenges relating to specificity of cause make use of the data and learning curve in gaining familiarity with them may not be trivial.

Recommendation: Procure and employ these data to answer specific questions of importance to CPSC that cannot be answered with CPSC data alone. Given the low costs, these data can be purchased without large budget allocations in advance.

Child Death Review Board Data

States throughout the United States have child death review boards that collect information on child deaths to determine causes of death, in part, and to identify instances of abuse, neglect, or homicide. Child death review teams have been active in the United States for more than 40 years. By 2001, all states had some form of child death review. There is some release and sharing of this information among states and with the federal government. This information could be a useful supplement to CPSC's MECAPS and other sources of data on deaths associated with consumer products.

The National Center for Fatality Review and Prevention (NCFRP) provides technical assistance and support to local and state child death review boards. They also maintain data from numerous state death reviews. As such, they may be privy to fatal incidents involving children and consumer products.

To date, the data have not been provided on a large scale in a systemic way to CPSC from a majority of the states. There are restrictions which vary by state on what can be shared by the review boards. Based on staff discussions with review boards, it appears that this creates a reluctance by the boards to even explore the possibility of sharing data, with efforts to date to obtain this information being unsuccessful, with most citing privacy laws and agreements. However, CPSC staff believes it may be possible to obtain these data in a legal

manner that is respectful of these agreements. Accordingly, CPSC has re-engaged with NCFRP to see whether at least partial data can be made available to CPSC with the right governing documents (*e.g.*, Memorandum of Understanding) in place.

Benefits

CPSC would obtain additional data on fatal consumer product-associated incidents involving children that may not be obtained via other methods.

Concerns

The states that could provide the data have demonstrated a reluctance to share the data (even in a manner that would satisfy any legal restrictions). Efforts to gain cooperation may not be fruitful.

Recommendation: CPSC staff should continue to pursue these agreements.

Poison Control Center Data

The National Poison Data System (NPDS) is the data warehouse for the nation's 55 poison control centers. Information is captured when people call the Poison Control Center after incidents of exposure, or possible exposure, have occurred. Depending upon the size and nature of the data requested, costs can range from \$10,000 to \$200,000 to obtain data. CPSC purchased the data in the past, but the agency ceased purchases, due to a low return on investment. Specifically, staff concluded that the data did not provide enough unique and specific information on poisonings (*i.e.*, beyond what is obtainable via NEISS) relative to the costs.

Benefits

CPSC would obtain additional data on incidents involving poisonings. Poison Control Centers make efforts to follow-up after calls to the center have been made. This permits determination of a final medical outcome in many cases, whereas NEISS only captures the outcome of an initial visit to the emergency department.

Concerns

Given CPSC's prior experience with the data, it seems likely a similar lack of utility relative to costs would be experienced.

Recommendation: Do not pursue at this time. Consider it as an option for projects with specific needs in the future where the unique information, such as follow-ups, would appear useful or necessary.

Exposure Survey Data

CPSC staff has previously made requests to the Commission for a long-term commitment of funds (likely around \$1 million annually) to collect exposure data directly from

consumers about the products they use. An exposure survey would permit more direct computations of risk for products under consideration by the Commission.

Benefits

CPSC could get direct, specific information on product ownership and use from an exposure survey. Using the results, CPSC staff could compute exposure and injury or fatality risk. This type of information helps inform the best understanding of risk, as it tells more than the number of products in the marketplace, adding critical information on how and how often they are used. For example, the durable nursery products exposure survey which CPSC reported in 2014 provided information not only on what fraction of US households owned specific products such as high chairs or booster seats, but also what fraction had them in use. Since a family may keep a high chair after their baby outgrows it and save it for when they have another child, this type of information helps staff understand correctly the frequency of use of products so that they correctly estimate the risk. If only a small fraction of products in homes are actually in use, the overall risk implied by the injuries is higher than it would appear if it were divided among all products owned.

Concerns

CPSC must consider numerous concerns associated with undertaking an exposure survey. The anticipated cost is not trivial and must be weighed against other Commission priorities. The costs include contracting costs of approximately \$1 million annually for an ongoing survey as well as staff resources of up to 1 FTE to gather inputs from technical staff, manage the project, administer the contract and monitor the contractor's progress. Only a limited set of products can be included as part of an exposure survey at a given time. Also, staff is aware that modern household surveys tend to achieve low response rates, which adversely impacts both cost and utility of the results.

Each survey project takes a long time to complete. The data collection cycles would include preparation of a questionnaire, OMB approval, and lengthy field periods (often 1 year or more). The OMB approval time would be reduced for a structured annual program which would involve follow-up approvals that are based on prior approval. CPSC could consider individual surveys for particular products or product classes on an ad hoc basis, however this approach would not benefit from the efficiencies of an annual program nor would they enjoy the benefits of economies of scale that associated with seeking owners of multiple different products at a time.

Recommendation: Consider whether the value proposition is sufficient for authorizing this ongoing expenditure. Staff may alternately recommend individual surveys as justified if the Commission does not support an ongoing expenditure.