Principles for Measuring the Impact of Voluntary Consensus Standards on Human Health and Safety

This page will be replaced with a Front Cover

Navigating through this document:

- Clicking in the table of contents takes you to the relevant page / section
- Opening the sidebar navigation pane in the pdf file also takes you to navigational links to the individual sections.

Printing the document: Note: This is a 45 page document.

Table of Contents

Acknowledgments

Sincere thanks are extended to all of the individuals and organizations listed below for providing technical input and/or other support associated with the development and promotion of this paper. It is because of their involvement and contributions that this document has been made possible. Special thanks go to those who submitted written content and comments for consideration, as well as those who provided editorial review.

The white paper is based on a consensus of those who actively participated in its development and does not necessarily reflect the views of the individuals or organizations listed below. The employment status and organizational affiliation of participants may have changed during the course of this project.

To be filled in.

Organization	Name of Individual(s)

[This page intentionally left blank]

Executive Summary

Measuring the impact of voluntary consensus standards on human health and safety is a complex undertaking. In 2021, staff of ULSE and CPSC approached ANSI to work collaboratively to better understand how standards and standards participation benefit society. Initially, a survey was conducted of the standardization community on efforts to address this issue. This led to the convening of a virtual workshop involving standards developing organizations, government agencies, and others, to share information and facilitate dialogue. Subsequently, a working group emerged from these activities and collaboratively developed this white paper.

What follows includes a discussion of:

- Existing efforts by the standardization community to measure standards' impact, and methodologies used.
- The goals of different stakeholders for measuring impact.
- Frameworks for conducting an impact assessment, including the different types of standards, and the importance of the voluntary standards adoption process.
- The components of measuring impact, including effectiveness and conformance models.
- The relationship between the effectiveness of a standard and conformance to a standard in achieving impact.
- Approaches to measuring the impact of standard, including challenges and quantifiers.
- Collecting and selecting data for measuring outcomes, with a discussion of assessment challenges, including confounding factors.

Ultimately, this paper seeks to illustrate the value of the work done by the standardization community and provide insight into how one can go about measuring the positive impact that standards have on human health and safety.

1. Introduction

The United States' Federal government has long recognized the value of standards developed by the U.S. private-sector led voluntary standardization system to support public policy objectives. The National Technology Transfer and Advancement Act (NTTAA), signed into law on March 7, 1996, directs Federal agencies to adopt voluntary consensus standards (VCS) developed by the private sector wherever possible in lieu of developing government unique standards.[1] OMB Circular A-119, last revised in January 2016, provides further direction on the Federal government's approach to engagement in standards development.[2] The circular notes, among other things, that

> *Executive Order 13563 ("Improving Regulation and Regulatory Review") emphasizes that the U.S. regulatory system "must protect public health, welfare, safety, and [the] environment while promoting economic growth, innovation, competitiveness, and job creation," and stresses the importance of public participation and careful consideration of both benefits and costs.*

Much has been written about the benefits of standards and standards participation.[3] But how does one measure the actual impact[4] that standards have on human health and safety? That is the focus of this white paper.

1.1 Background

The U.S. Consumer Product Safety Commission (CPSC) "works to save lives and keep families safe by reducing the unreasonable risk of injuries and deaths associated with consumer products."[5] Under the CPSC's authorizing legislation,

> *(b)(1) The Commission shall rely upon voluntary consumer product safety standards rather than promulgate a consumer product safety standard prescribing requirements described in subsection (a) whenever compliance with such voluntary standards would eliminate or adequately reduce the risk of injury addressed and it is likely that there will be substantial compliance with such voluntary standards.[6] SEC. 7. [15 U.S.C. §2056]*

---

[1] NTTAA, Public Law 104-113, accessed May 1, 2023, https://www.govinfo.gov/content/pkg/PLAW-104publ113/pdf/PLAW-104publ113.pdf

[2] OMB Circular A-119: Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities, accessed May 1, 2023 https://www.nist.gov/system/files/revised_circular_a-119_as_of_01-22-2016.pdf

[3] See, for example, Standards Boost Business, accessed May 1, 2023. Also, Allen, R. H., & Sriram, R. D. (2000). The role of standards in innovation. *Technological Forecasting and Social Change, 64*(2-3), 171-181.

[4] For purposes of this paper, impact means the actual change(s) (both positive and negative) that the standard has produced on the impacted population. These changes may generally be categorized as mortality, morbidity, or property loss. See definitions in the Appendix.

[5] About CPSC, accessed May 1, 2023 https://www.cpsc.gov/About-CPSC

[6] Consumer Product Safety Act (Public Law 92-573; 86 Stat. 1207, Oct. 27, 1972) accessed May 1, 2023 https://www.cpsc.gov/s3fs-public/pdfs/blk_media_cpsa.pdf?epslanguage=en

Accordingly, CPSC provides accident and injury data and otherwise supports efforts to develop voluntary standards, working with standards developing organizations (SDOs) and industry. CPSC staff's initial interest in the topic of this white paper was to try to gauge the overall benefit that agency participation in standards provides to the American people.

UL Standards & Engagement (ULSE) works to "develop and publish consensus standards that help guide the safety, performance, and sustainability of new and evolving products, technologies, and services that range from household appliances, smoke alarms, and batteries to building materials, cybersecurity, and autonomous vehicles."[7] ULSE asked its technical committee members why they participate in standards development. One of the most common responses was that members felt they were doing a good thing for society. That the mission of undertaking standards development for security, sustainability, and safety, is something that they personally get gratification from. ULSE realized that one way to attract more volunteers is to be able to measure the impact of standards and report back that it is leading to better outcomes for society. Such metrics will help people to get involved and stay involved.

The American National Standards Institute (ANSI) serves as administrator and coordinator of the U.S. private sector system of voluntary standardization. ANSI has a successful track record serving as a neutral facilitator to convene stakeholders from the public and private sectors to address national standardization priorities. CPSC and ULSE reached out to ANSI to assist with obtaining perspectives from the standardization community on best practices and methodologies related to measuring the impact of voluntary standards on human health and safety.

After gathering data via an initial survey, ANSI with assistance from ULSE and CPSC staff convened a Virtual Workshop, "Measuring the Impact of Voluntary Consensus Standards on Human Health and Safety," on Thursday, October 28, 2021 (Link to full report). This initial meeting brought together a diverse group of speakers representing trade associations and professional societies, SDOs, government regulators, and consumers. The goal was to initiate a dialogue, build upon existing work, and fuel future collaboration to develop common measurement methodologies.

After that meeting, a Working Group was convened among those who expressed interest in continuing the discussion with a view toward developing a white paper describing principles for measuring the impact of standards. This document represents the outcome of that Working Group effort, after almost two years of meetings and collaboration.

1.2 Structure of this Document

Section 2 describes existing efforts by SDOs, other standards organizations, and U.S. federal government agencies to measure standards' impact, as well as considerations that are important for consumers and labor organizations.

[7] ULSE accessed May 1, 2023 https://ulse.org/

Section 3 provides more insight into the goals of different stakeholder groups (SDOs, U.S. government agencies, industry associations, advocacy organizations) on measuring standards' impact.

Section 4 discusses frameworks for planning and executing an impact assessment. This includes the applicability of different types of standards to impact assessment, the voluntary standards adoption model, how impact varies during the lifecycle of a standard, and how to conduct an impact assessment.

Section 5 describes impact components, including how effectiveness and conformance are related to impact and approaches to measurement.

Section 6 provides guidance on measuring impact. This includes selecting or collecting data, establishing a baseline, determining the outcome, form of analysis, and confounding factors.

A summary of Conclusions and an Appendix defining key terms complete the document.

2. Existing Efforts to Measure Impact

This section recaps the activities and learnings shared during the October 28, 2021 Virtual Workshop and includes updates since that time, and plans for future work. It also includes additional projects and perspectives not discussed at the workshop that were provided by working group members. It describes issues that have arisen (e.g., obstacles to accessing data) and lessons learned, including success stories.

Working group members were asked to present:

- What initiatives (studies, publications, surveys, etc.) has the organization/agency undertaken to measure standards' impact;
- To explain their approach, whether it was effectiveness,[8] conformance,[9] or a mix of the two; and
- Key findings of these initiatives and next steps.

2.1 Standards Developing Organizations (SDOs)

2.1.1 UL Standards & Engagement (ULSE)

In March of 2021, ULSE launched an initiative to measure the impact of its standards on public health and safety. Three exemplar standards were considered for evaluation: *ANSI/CAN/UL 325 Standard for Door, Drapery, Gate, Louver, and Window Operators and Systems; ANSI/UL 859, Standard for Household Electric Personal Grooming Appliances;* and *ANSI/CAN/UL 2272, Standard for Electrical Systems for Personal E-Mobility Devices.* ANSI/CAN/UL 325 was selected for a pilot study. Data on entrapment incidents was collected and examined to render a judgment on the efficacy of the inherent and external entrapment protection requirements in UL 325 for residential garage door operating systems in the United States., Incident data examined included CPSC's National Injury Information Clearinghouse (Clearinghouse) and National Electronic Injury Surveillance System (NEISS). NEISS provides a nationally representative probability sample of injuries associated with consumer product injuries seen in hospital emergency departments. The Clearinghouse is an amalgamation of sources including death certificates, coroner and medical examiner reports, investigations from the CPSC and other governmental agencies, and incidents reports received from consumers and manufacturers to the CPSC.

Early on, ULSE found shortcomings in the approach of focusing exclusively on the existing data on injuries with consumer products. The first constraint encountered was the limited coverage of injuries occurring with consumer products. NEISS does not include injuries treated at other medical providers including primary care physicians, urgent care facilities, and independent medical clinics, nor does it

[8] For the purposes of this paper, effectiveness is the ability of the standard to produce a desired change in the impacted population. See definitions in the Appendix.

[9] For purposes of this paper, conformance is a composite factor characterizing both the quantity of [products] meeting the standard, and the degree of the adherence of these [products] to the standard.

include minor injuries not requiring medical attention. The Clearinghouse only captures information on injuries if they are mentioned in the initial report, which is not uniformly the case.

The second constraint involved the suitability of raw injury counts for assessing impacts on safety. Without the knowledge of the number of a product in the marketplace (and the changes in those numbers), how could one know whether a decrease in injuries from a consumer product was due to improvements in the safety of the product and not simply a reflection of fewer units of the product on the market and therefore fewer injuries? Any change in the numbers of injuries from a consumer product following the publication of a standard without any knowledge of the actual "exposure" could give a false impression as to any actual changes in the safety of that product.

The final limitation was the difficulty in credibly inferring any improvements in the safety of a product to the standard and not some other factor(s). There was no ability in the study to account for, much less control, the myriad of other possible contributing factors to a reduction in the injuries.

To overcome these weaknesses, the assessment process incorporated some of the features of the Centers for Disease Control *Framework for Program Evaluation in Public Health*. Elements to be incorporated include:

- engaging stakeholders from outside ULSE into the assessment process,
- developing an underlying logic model with the assistance of these stakeholders to illuminate the relationship between publication of a standard and reductions in injuries with covered products,
- and measuring leading indicators based on the logic model along the path from publication of a standard to a reduction in injuries.

Despite the restrictions of the available data and the limitations of the study design, a credible determination that UL 325 has had a positive impact on public health and safety through the mitigation of the risk of entrapment by electronically operated residential garage door operating systems was determined.

2.1.2 CSA Group

The mission of CSA Group's standards development organization is to enhance the lives of Canadians through the advancement of standards in the public and private sectors. In May 2021, CSA Group (CSA) initiated "Project Confirmation" aimed at studying where CSA standards are referenced by industry and governments, and the health, safety, economic, social, and environmental impacts of these standards. Project Confirmation is aimed at answering two fundamental mission fulfillment questions:

- Are the lives of Canadians being enhanced through the use of CSA standards?
- If so, in what specific ways are CSA standards being impactful?

At project initiation, CSA studied about 10 percent of its estimated standards portfolio, and focused on the gas and health care sectors only. Subsequently, the project studied the entire portfolio of CSA

standards – just over 3,200 standards. The project catalogued where the standards are referenced publicly by industry and government (federal, provincial, territorial, and municipal). The data provides insight into the nature of these references: Are standards being mandated or are they voluntary? Do the references imply the use of the latest version of a standard (ambulatory) or refer to a fixed version (static)? Who references the standards?

The data is ready to be analyzed, and once CSA knows how many standards are being referenced, who uses them, and how they're being used, it will measure their health, safety, economic, social, and environmental impacts. Once the methodology for this work has been put in place, the intent is to track and measure utilization and impact trends over time.

Why is Project Confirmation vital ?

CSA is 103-years young. It is an organization that has pioneered standardization in key sectors of the economy, from the first set of standards for railroad bridges, to cutting-edge standards that speak to present and future needs, such as electric vehicles, digital technology, bioeconomy, climate change, etc. With a portfolio of over 3,200 standards, it is imperative that organizational impacts on stakeholders are measured and communicated regularly.

CSA has many stakeholders in different sectors, and so the impact of standards on them will vary. Questions that arise include: Are CSA standards making it easier to choose and compare products by increasing interoperability? Are they making products safer and more durable for consumers? Are CSA standards providing certainty in fields such as construction and manufacturing? Are they helping government legislators and regulators when they begin to think about mandatory requirements? How do CSA's environmental standards help stakeholders reach their environmental stewardship goals? Do they help to reduce greenhouse gas emissions and pollutants?

From an organizational perspective, knowing where CSA standards are utilized, and what impact they're having, will help CSA allocate resources to increase utilization and update standards to make them even more impactful. One of the data points that CSA is seeking is the time lapse between the publication of a standard and the first reference of that standard by governments or industry. If the time is reasonably short, that would indicate that the standard is deemed to be of value. However, if it is a long period of time, that may warrant investigation into why, and perhaps committing additional resources for advocacy and government relations.

This project concluded in 2022, following which CSA will have a tool that will enable this work to continue over time, so that it can continually track the utilization and impact of its growing portfolio of standards.

2.1.3 American Society of Safety Professionals (ASSP)

ASSP has heard a lot from occupational safety and health (OSH) professionals who ask: What's the point of voluntary standards? Can you prove they are effective in improving workplace safety? Why should we

use consensus standards when we have government regulations? In response ASSP has investigated the use of management systems standards. In a white paper titled *The Return on Investment for Safety, Health, and Environmental (OSH) Management Programs*, ASSP provides evidence of the positive financial benefits to implementing OSH management systems. Examining both the direct and indirect costs of workplace injuries, including worker's compensation claims, liability and legal expenses, decreased productivity, and lower employee morale, ASSP makes the case that organizations with OSH management systems have experienced dramatically reduced costs associated with workplace injuries. (CUs)

ASSP serves as secretariat of the Z10 Committee responsible for the Z10 Occupational Health and Safety Management Systems Standard. It also serves as the U.S. TAG Administrator for the ISO TC 283 Committee responsible for the ISO 45001 OHSMS Standard. In September 2021, the U.S. Mine Safety and Health Administration (MSHA) published a proposed rule on surface mobile equipment that cited the Z10 and ISO 45001 standards among other safety program guidance materials that it reviewed. Based on its review of these materials, other best practices, and comments, MSHA "concluded that developing and implementing a written safety program for surface mobile equipment at mines would contribute to advancing miners' safety and health." ASSP would like to see more future public and private sector use of the standards to provide implementation data and show their value.

Another initiative that ASSP has taken the lead on is the use of leading indicators (e.g., measures of efforts to mitigate workplace injuries) in the management of OSH. The ASSP Z16 committee, with about 60 organizations, has published a standard that defines the requirements of a set of leading indicators that can be used by organizations to effectively measure their safety and health performance: *ANSI/ASSP Z16.1-2022 Safety and Health Metrics and Performance Measures*. Traditionally, the regulatory-driven approach to OSH has focused exclusively on lagging indicators (e.g., number of people injured, workers compensation claims, etc.). ASSP believes that by including leading indicators into their assessment process, employers will be better able to manage their injury prevention efforts, ultimately reducing workplace injuries. Key takeaways from ASSP's work with Z16 and the incorporation of leading indicators into measures of safety and health performance are: 1) prevention efforts in implementing organizations are improved 2) costs (both direct and indirect) associated with workplace injuries are reduced, and 3) worker participation in OSH initiatives is increased.

2.1.4 ASTM International

In 2012, CPSC issued a safety alert that liquid laundry packets (LLPs) should be kept away from children after incident reports associated with LLPs began to appear in the surveillance data maintained by the CPSC. The main hazards were ingestions that could lead to fatalities and ocular injuries that require medical assistance. In 2013, CPSC Chair Inez Tenenbaum requested voluntary action by industry. Within months, ASTM International, a voluntary standards development organization, convened a group of stakeholders to address the hazards associated with LLPs through the development of a voluntary consensus standard. ASTM formed subcommittee F15.71 on Liquid Laundry Packets, which brought together manufacturers, consumer groups, medical professionals, regulators, testing labs and others.

The ASTM meetings resulted in the publication of a new voluntary standard, F3159-15e1 *Standard Safety Specification for Liquid Laundry Packets*, in October 2015. The standard addresses hazards such as packaging, labeling, and taste/dissolution properties of LLPs.

By December 2016, industry reported that these voluntary safety measures were fully implemented, with nearly all LLPs available for sale to consumers becoming compliant with ASTM F3159-15e1. The subcommittee aimed to not only focus on conformance to the standard, but also on the effectiveness and making sure the standard made an impact on safety.

Much of the work in evaluating the impact of the standard on safety took place through the coordination of several ASTM tasks groups to evaluate the data and injury trends, scope of the standard, and additional measures to consider. The ASTM data task group sought to monitor injuries associated with LLPs before, during, and after implementation of the standard. The task group used two main data sources to monitor the impacts: The Rocky Mountain Poison & Drug Safety (RMPDS) National Poison Data System (NPDS) and CPSC data obtained through NEISS.

During the initial review of incident data for the post implementation period, the subcommittee was committed to ensuring comprehensive data analysis, as much as possible, to determine the standard's impact. Initially, the subcommittee was reviewing data that compared the number of incidents to the population rate; however, it was later agreed to compare the number of incidents against the number of LLPs sold since the number of LLPs in the market kept increasing and this denominator influenced the data. As the subcommittee continues to evaluate the reported incidents, they continue to consider the sales-adjusted rate and other ways to interpret the data against reported incidents.

One analysis, based on NPDS data, showed that exposures to LLP hazards increased through the transition period but began decreasing in 2017. The RMPDS, which analyzed the data, also found a significant drop in clinically significant outcomes, including severe medical outcomes (i.e., life-threatening effects or death) in the population and sales-adjusted rates of exposures over the study period. The center also noted a drop in observed exposures for children under two.

The second analysis was conducted by the CPSC using NEISS data. The agency did not find improvements in absolute number of estimated emergency department visits between the baseline and post-implementation period but found significant declines in hospitalization and sales-adjusted emergency department visits.

The ASTM F15.71 subcommittee continues to meet on a regular basis to review incident data against requirements in the standard for the purposes of determining if the standard is addressing known hazard patterns. Since the original publication date of F3159-15e1 in 2015, the subcommittee reviewed five reports prepared by the CPSC for the post-implementation period and annual reports from RMPDS. Additionally, the subcommittee stays on top of published research literature relating to incidents. During the COVID-19 pandemic, CPSC also conducted a study on exposures and incidents of household products during the pandemic when many people spent more time in their homes. This data was shared with and reviewed by the subcommittee. The 2015 standard was reaffirmed as written in 2022.

A valuable outcome of the ASTM F15.71 standards work is the creation of joint training programs offered in collaboration between ASTM and the International Consumer Product Health and Safety Organization (ICPHSO). Topics covered in these training programs include information and education campaigns for consumer product safety hazards, warnings, and age grading for toys and children's products.

ASTM International rules require that standards are reviewed for ballot at least every five years. If changes in reported incidents lead to a need to revise or update the standard, the subcommittee can do so at any time through ASTM's balloting process

2.2 Other Standards Organizations

2.2.1 Standards Council of Canada (SCC)

What initiatives (studies, publications, surveys, etc.) has the organization/agency undertaken to measure standards' impact?

- SCC conducted two studies looking at the impact of standards on unintentional fatalities.
- The studies did not look at individual standards but rather at participation in ISO TCs as a proxy for standardization activity.
- The first study found an association between TC participation and unintentional fatalities, such that countries that were more involved in standards development had fewer people dying as a result of unintentional injuries.
- However, a follow up study found that the relationship between TC participation and unintentional fatalities was driven by the results for men. TC participation is associated with a reduction in unintentional fatalities for men, however, there is no such relationship for women. This study highlights the importance of considering gender in standards development.

Is the focus on effectiveness, conformance, or a mix of the two?

- As a national standards body (NSB) that does not develop standards, SCC studies have tended to focus more on "system level data," e.g., TC participation, the number of active standards in the SCC catalogue, etc.
- SCC has some research looking at individual standards that it has commissioned, but to date it has not had sufficient data to reliably assess the impact of these standards.

What were the key findings of these initiatives (if known at this time)? Issues that arose, obstacles overcome, useful outcomes? (Include any URLs to published studies.)

- By using cross-country data, SCC was able to show the relationship between standards and health and safety. However, SCC also showed that the relationship is influenced by gender. Gender is not the only consideration. Additional research is needed to address other demographic variables and intersectionality more fully.

What next steps, if any, does the organization plan to take as a result of these initiatives (or as a continuation of them if still in progress)?

- The gender research has been used to highlight the importance of gender-responsive standards.
- SCC is co-leading the ISO/IEC JSAG on gender-responsive standards.
- SCC is the Chair of the United Nations Economic Commission for Europe (UNECE) Team of Specialists on Gender-Responsive Standards which has released guidelines for how to develop gender-responsive standards.
- SCC also has corporate objectives to improve gender-responsiveness in standards development.

2.3 Industry Associations

2.3.1 International Safety Equipment Association (ISEA)

ISEA is a trade association for safety equipment manufacturers, distributors, and testing laboratories. In addition to policy advocacy, ISEA is an ANSI-accredited standards developer and works to ensure that performance standards for safety products are adopted by reference in regulation. ISEA has developed a number of standards for personal protective equipment (PPE) and body safety equipment. ISEA also provides a voice for the safety equipment industry in other SDOs that develop standards for or that utilize its products.

Standards for test methods and predictive performance standards for safety equipment products are building blocks for safety management system standards and for regulation. Measuring the impact of these is important. Without predictive performance, there is a huge decrease in adoption of safety equipment in the field, in regulation, and in safety management systems. Thus, providing minimum performance standards for safety equipment is a bedrock for safety.

ISEA measures the impact of safety equipment products to mitigate injury and illness. It may also measure the uptake of safety equipment into regulation or the number of workers protected. ISEA undertook an effort to quantify the economic impact of the safety equipment industry using the number of jobs created as a measure. However, ISEA's leadership felt that the economic impact of the industry was secondary to its impact on protection. So, ISEA also measured the total number of employees in the U.S. workforce who are protected by standardized safety equipment. ISEA now has data at the U.S. federal, state, and local level by broad sector. It plans to replicate this data year after year with the goal of producing trend lines as new safety standards are promulgated or adopted into regulation to see how adoption of safety equipment changes over time.

2.3.2 Association of Home Appliance Manufacturers (AHAM)

What initiatives (studies, publications, surveys, etc.) has the organization/agency undertaken to measure standards' impact?

- AHAM has done a SWOT (Strength-Weakness-Opportunities and Threats) analysis of their standards. This was mostly an internal assessment but the results were shared with AHAM members for concurrence and additions of things that may have been missed.

Is the focus on effectiveness, conformance, or a mix of the two?

- Effectiveness and relevance to consumers

What were the key findings of these initiatives (if known at this time)? Issues that arose, obstacles overcome, useful outcomes? (Include any URLs to published studies

- It was a good exercise to make AHAM aware of the actual usage of their standards, the benefits they provide, the future risks to the standards, and where there may be global harmonization opportunities.

What next steps, if any, does the organization plan to take as a result of these initiatives (or as a continuation of them if still in progress)?

AHAM will revisit the SWOT every couple of years for updates or addition of new information. AHAM is considering developing an assessment tool to measure the effectiveness (impact) of its standards work.

2.4. U.S. Federal Government Agencies

2.4.1 U.S. Environmental Protection Agency (EPA)

EPA has a continuing commitment to using and participating in private sector standards as directed by the NTTAA and other federal policies. Use of standards in regulation, voluntary programs, research, and other activities helps the agency achieve robust engagement with industry, academia, non-governmental organizations (NGOs), and others. It also helps to harness partnerships that enhance public trust and leads to durable policies that enjoy broad consensus and buy-in. Learn more at www.epa.gov/vcs.

With regard to EPA's regulatory activities, EPA uses several types of economic analyses to inform many aspects of regulations. Key among these analyses is benefit-cost analysis (BCA), which estimates the expected social costs, social benefits, and then the net benefits of various regulatory options. BCA at the EPA is guided by the Agency's Guidelines for Preparing Economic Analyses, as well as federal guidance that applies to all regulatory agencies (e.g., OMB Circular A-4). BCA is designed to assess the efficiency of regulatory standards by characterizing, to the extent practicable, all costs and benefits. If expected benefits exceed costs, then the standard is likely to be a move toward economic efficiency. EPA often uses additional tools such as cost-effectiveness analysis to provide additional insight into the merits or effectiveness of regulations.

Specific findings of BCA depend on the nature of the regulatory standard and the availability of data and methods to fully characterize both benefits and costs. In some cases, costs and benefits can be

quantified well and conclusions about net benefits are relatively clear. For example, EPA's analyses of the costs and benefits of the Clean Air Act shows that benefits are much greater than costs, i.e., that net benefits are large and positive.[10] In other cases, limited data and methods make it difficult to quantify costs and, perhaps more frequently, benefits. EPA's 2017 proposed regulation on paint strippers provides one clear example.[11] The proposed rule would regulate the chemical N-Methyl-2-Pyrrolidone, or NMP, for certain uses. NMP is known to have adverse human health impacts on fertility and development, but current risk assessment methods do not provide a dose-response function for these effects. Without such a function, the benefits of reduced exposure to NMP cannot be quantified in the benefit-cost analysis. It is also often challenging to quantitatively estimate the ecosystem benefits of environmental regulatory standards.

When EPA is able to quantify expected changes to health and health risks that result from regulations, BCA at EPA relies upon two primary methods for estimating dollar values for those changes: willingness to pay (WTP) and cost of illness (COI). WTP is the core concept for valuing benefits within BCA, and it is a comprehensive measure of what people are willing to give up, in terms of income or wealth, to obtain the expected health improvements from regulatory standards. For health, referred to as "morbidity" in EPA's analyses, WTP reflects four components of value: avoided medical expenses, avoided loss of productive time, avoided defensive expenditures (i.e., spending to protect oneself from whatever the risks are), and avoided pain, suffering, and discomfort. Reduced risk of premature mortality, a benefit from many regulatory standards, is usually expressed as the value of a statistical life (VSL).

It can be challenging to estimate WTP. For most health effects, WTP cannot be estimated directly from market outcomes. Instead, economists look at related markets where people undertake some action that reveals their preference for risk reduction. For example, consumer expenditures on protective goods (e.g., water filters) may shed light on what people are willing to pay to reduce risks to health and safety. A key source for some WTP values is hedonic wage (or wage-risk) studies. The underlying logic is that wages are driven by many factors, one of which is the risks faced in the workplace, e.g., the risk of fatal injury and the risk of non-fatal injury. Careful analysis of how wages differ systematically with workplace risks provides insight into workers' willingness to trade off higher compensation for higher risks. Such studies are widely used for VSL estimates by federal agencies. Another source for WTP estimates is hypothetical markets, where people's stated preferences for risk reduction are captured – very carefully and systematically – through survey methods.

While WTP is the preferred measure of health and safety benefits in BCA, it is often not available. A more limited measure is cost-of-illness (COI) which is usually composed of two factors: medical expenses avoided from a given injury, and the value of lost time. In principle, the value of lost time would include the value of lost work time, lost household production (e.g., taking care of the home, children, etc.) and lost leisure time. In practice, COI estimates may often only include lost work time. COI is more limited

[10] US EPA. 2011. The Benefits and Costs of the Clean Air Act from 1990 to 2020. https://www.epa.gov/sites/default/files/2015-07/documents/fullreport_rev_a.pdf

[11] https://www.regulations.gov/docket/EPA-HQ-OPPT-2016-0231

than WTP in large part because it does not include the value of avoided pain, suffering, and discomfort, which may be substantial.

EPA continues to perform BCA for proposed and final regulations and has ongoing research to address many of the challenges to estimating both costs and benefits. This includes research to estimate WTP for more health effects, application of methods to improve the ability to quantify expected changes in health outcomes from different standards, and the application of more sophisticated models for both costs and benefits.

2.4.2 U.S. Food and Drug Administration (FDA)

The FDA's Standards and Conformity Assessment Program (S-CAP) leads and promotes the use of VCS in support of its regulatory mission to ensure patients access to safe and effective medical devices. Part of the FDA's Center for Devices and Radiological Health (CDRH), S-CAP conducts this work through a robust standards recognition program that relies upon international consensus standards and builds upon the Total Standards Life Cycle (TSLC), through which the impact of standards is monitored and analyzed.

S-CAP staff also direct the participation of hundreds of CDRH experts in more than six hundred SDO committees and working groups and manage systems that track the FDA's contributions to these standards. Standards that receive FDA recognition are available in a publicly-accessible database that features a trove of resources to encourage manufacturers to cite recognized standards in device regulatory submissions.

In order to advance the use of standards, S-CAP analyzes their use in device submissions. The most recent research, conducted in 2015, measured how often FDA-recognized standards were cited and whether those citations were correctly accompanied by a declaration of conformity (DOC) and the appropriate supplemental documentation, versus "General Use" of a non-recognized standard. While manufacturers may use any standard under "General Use," the FDA strongly encourages the use of recognized standards for their potential to streamline the regulatory review of the device.

In this 2015 study, the FDA surveyed a sample of submissions across all twenty medical device specialties from 09/01/2014 through 09/01/2015. FDA observed that the citation of VCS in premarket submissions is high, with an average of seven standards in each submission. Forty-seven percent of the submissions used recognized standards with a DOC, an encouraging sign. FDA also identified the ten most commonly used recognized standards in the sampled data. The results of this analysis highlighted some confusion about the benefits of using a recognized standard with a DOC, as well as errors such as citing the wrong version of a standard or deviating in the testing method specified but still treating these as recognized standards. Based upon this research, the FDA published *Appropriate Use of Voluntary Consensus Standards in Premarket Submissions for Medical Devices* guidance in 2018 to advance the correct use of standards.

S-CAP is in the midst of an update to the 2015 research. This new study explores and measures inconsistencies in how DOCs and supplemental documentation are submitted by manufacturers and

how they are reviewed by the FDA when consensus standards are cited. The goal for this project is to develop tools and resources, e.g., checklists and templates, to facilitate the use of standards and ultimately to make the review process more efficient.

Finally, the S-CAP team has, after significant analysis and consideration based upon its previous research, conceptualized CDRH's first conformity assessment scheme. Designed to streamline the conformity assessment aspects of device review, the *Accreditation Scheme for Conformity Assessment*, or ASCA, is in its early implementation phase and beginning to show results, which S-CAP will evaluate in the coming months. The ASCA program fosters organic collaboration between standards developers, conformity assessment bodies and the FDA, to put standards to work more effectively, and encourage their use by manufacturers to demonstrate safety and effectiveness.

2.4.3 Centers for Disease Control and Prevention (CDC)

In 1999, CDC published *Framework for Program Evaluation in Public Health* (Framework). Since its release, the Framework has been the playbook for CDC program evaluations, allowing them to become more contextually grounded, participatory, appropriately designed, and effectively implemented. Central to the Framework is the idea that a good evaluation collects accurate evidence to draw valid conclusions about the program's effectiveness and provides results that can be used to improve the program. While the relevance of the Framework to assessing the impact of voluntary consensus safety standards may not be initially clear, both voluntary consensus safety standards and public health programs share the goal of seeking to prevent injury, illness, and death. Furthermore, both require the participation of external stakeholders to successfully achieve their intended safety and health outcomes.

The Framework is defined by six sequential steps and four sets of guiding "standards" for conducting sound evaluations of public health programs (Figure 1). The six steps include: engaging stakeholders, describing the program, focusing the evaluation, gathering credible evidence, justifying conclusions, and ensuring use of evaluation findings and sharing lessons learned. The four sets of guiding "standards" consist of utility, feasibility, propriety (ethically conducted), and accuracy.

Figure 1: CDC Evaluation Framework
(from U.S. Department of Health and Human Services & Centers for Disease Control and Prevention (2011). *Introduction to program evaluation for public health programs. A Self-Study Guide*, 2005.)

The Framework places special emphasis on setting the appropriate focus for the evaluation (Step 3) through engaging stakeholders in the evaluation (Step 1) and clearly describing the program to be evaluated (Step 2). At the heart of the program description is the logic model, providing a graphical depiction of the relationship between program activities and intended effects. From the logic model the appropriate measurable outputs from the program activities can be identified and examined to reach valid conclusions.

The Framework provides a blueprint to ensure that evaluations of public health programs will meet the diverse needs of stakeholders including assessing and documenting program implementation, achieving desired outcomes through efficient activities, and identifying actions based on evaluation results to increase the impact of the programs evaluated. While there are some fundamental differences between evaluating public health programs and assessing the health and safety impacts of VCS, there are also important lessons that the CDC has learned from its decades-long journey with the Framework that can be applied to improve other impact assessment processes.

2.4.4 U.S. Consumer Product Safety Commission (CPSC)

In March 2004, CPSC docketed a petition from the National Association of State Fire Marshals (NASFM)[12] requesting the Commission mandate fire safety standards for candle products based on, at a minimum, the requirements contained within the ASTM *Provisional Specification for Fire Safety for Candles* (PS59-02), published by ASTM in 2002. In July 2006, the Commission voted to defer a decision on the petition and directed staff to continue working with ASTM to develop standards for candle products and to provide periodic status updates on the developments of the standards to the Commission. In August

[12] CP 04-1/HP 04-1, Petition for improved candle fire safety, National Association of State Fire Marshals, February 10, 2004

2014, the Commission voted to deny the petition as data showed the standards were effective and widely adopted.[13]

ASTM published a performance standard for candle safety, ASTM F2417 in 2004 then later published a performance standard for candle accessory safety, ASTM F2601 in 2007. Both standards have been revised since initial publication, with the most recent versions being ASTM F2417-17 and ASTM F2601-18. CPSC and NFPA data showed that residential fires, deaths, and injuries associated with candles all generally increased through the 1990s peaking in the early 2000s followed by a decline into the 2010s. A confounding factor in this analysis was a decrease in candle sales between the early 2000s and early 2010s. However, residential fires decreased 58% while candle sales only decreased by 35% over a similar time period. The petitioner, NASFM, believed the ASTM candle standards would be effective, and data suggested that to be true, but the petitioner worried that industry compliance would be low.

In comments on the petition, the National Candle Association (NCA) maintained that its members produce candles and candle products "in accordance with recognized industry standards and practices." Because its members represent 90 percent of candles manufactured in the United States, the NCA argued that U.S. production is in substantial conformance with the current ASTM standards. Likewise, the Household & Commercial Products Association (HCPA),[14] commented that its members, who include "most of the major candle manufacturers and marketers in the United States [are] in compliance with the current ASTM standards." Additionally, several of the largest candle manufacturers and three of the largest candle retailers actively participate in the ASTM subcommittee that maintains the candle standards. This information suggested substantial industry compliance to the ASTM standards.

The CPSC must defer to voluntary standards when standards exist, the standards are effective, and the standards are widely adopted by industry. CPSC staff determined that the ASTM standards were effective, and industry largely produced compliant candles. However, the industry did also experience a decrease in sales over the same time period not allowing for a full analytical analysis of the overall impact.

2.5 Consumers

While this paper has not identified activity by consumer organizations to measure the impact of standards, conversations with consumer advocates revealed a preference for standards development that is characterized by the following attributes:

- A strong and transparent standards development process has adequate participation by consumers and regulators.

[13] https://www.cpsc.gov/s3fs-public/pdfs/foia_PetitionCP04-1andHP04-1RequestsforFireSafetyStandardforCandlesandCandleAccessories.pdf

[14] Previously known as the Consumer Specialty Products Association (CSPA)

- There is the potential for regulatory oversight if warranted.
- The process should not be wholly controlled by the regulated industry.
- The standard must effectively address the hazard(s) at issue.
- The standard has to be widely used and accepted.

The *ANSI Essential Requirements: Due process requirements for American National Standards* (*ANSI Essential Requirements*) and associated guidance documents [15] recognize both the role that consumers play as key stakeholders and the critical nature of openness, lack of dominance, and a balance of interests within a consensus body. ANSI has a full-time staff member who works to assist standards developers find, train, and retain consumers for participation on committees.

2.6 Labor

For 30 years, the American Federal of Labor and Congress of Industrial Organizations (AFL-CIO) has produced an annual report on the state of safety and health protections for America's workers: *Death on the Job, The Toll of Neglect, A National and State-by-State Profile of Worker Safety and Health in the United State (30th edition, May 2021)*. The report notes that, in the 50 years since the Occupational Safety and Health Act went into effect, there have been significant improvements in worker protections with many safety regulations having been promulgated by federal agencies and hundreds of thousands of lives saved. The report, among other things, calls on OSHA to promulgate standards to protect workers from COVID-19 and other infectious diseases, and to protect healthcare and social service workers from workplace violence.

[15] See, for example, *Guidance on Balance and Outreach within the American National Standards (ANS) Process* (ExSC_013_2022) and *Guidance on Lack of Dominance within the American National Standards (ANS) Process* (ExSC_013_2021).

3. Stakeholder Goals for Measuring Impact

While standards are written for the overall benefit of society, and in this case the improvement of health and safety; each stakeholder has their own goals, perspectives, and motivations related to impact and its measurement. This section articulates the goals of various stakeholder groups (e.g., SDOs, U.S. government agencies, industry associations, and advocacy organizations).

3.1 Standards Development Organizations (SDOs)

The purpose of an SDO is to promulgate standards that are within the scope of its authorization and/or accreditation, and that will be used and useful. Measuring the impact of a standard's content provides evidence that the standard is used and useful and the SDO can assess if the content of the standard achieves the stated purpose of the standard.

SDOs can use impact measurement to take actions to increase the beneficial impacts of the standard(s). When a standard's impact is less than anticipated, the SDO can use this information to further revise and improve the standard. Additionally, this impact assessment method may be used to understand content within a standard that does not provide commensurate improvements in health and safety.

By measuring the impact of [product][16] standards, SDOs can demonstrate that their [product] standards contribute to improved health and safety; this can encourage continued funding of, and participation in, the standards efforts.

Impact measurement will help ensure that the limited resources contributing to a standard (SDO staff, technical committee members, etc.) are aligned to the standards/opportunities that have a beneficial impact on human health and safety.

3.2 Industry Associations

One of the goals of industry associations is to promote the use of the industry's [products], informing stakeholders about the industry's benefits. By measuring the impact of standards, industry associations can demonstrate that their [products] contribute to improved health and safety.

The motivations of industry associations in participating in standards development and impact assessment may include the establishment of baseline safety and performance criteria for the industry, protecting the industry workforce, managing the industry's costs, and ensuring social benefit from the industry's [products].

[16] For the purposes of this document, [product] with brackets means the output of an organization that can be produced without any transaction taking place between the organization and the customer or end user [product]. This includes physical products, processes, services, methods, personnel, and other intangible assets subject to the standard. See definitions in the Appendix.

Some industry associations are also SDOs; for these industry associations, their motivations will be aligned with those of SDOs.

3.3 U.S. Federal Government Agencies

Use of standards, and participation in their development can support government agencies in their regulatory, policy, procurement, and other activities. In fact, the NTTAA: (1) directs agencies to adopt VCS wherever possible (avoiding development of unique government standards), and (2) promotes agency participation on standards bodies. Government use and participation in standards benefits agencies in many ways, including:

- Robust engagement and partnership with diverse stakeholders
- Durable policies, broad consensus and buy-in
- Efficient and effective use of public resources
- Compliance with international trade laws and obligations
- Scientific exchange to spur innovation and to advance agency mission
- Development of agency staff technical and leadership skills

Longstanding Executive Orders require agencies to conduct a regulatory impact analysis for economically significant rules, including those that incorporate standards by reference[17]. In evaluating whether to use a standard, agencies are directed to consider (among other factors):

> "…the level of protection the standard provides or is expected to provide for public health, welfare, safety, and the environment."

> "…the costs and benefits to the Federal government and the regulated public of the agency developing its own standard…"

> "…the prevalence of the use of the standard in the national and international marketplaces."

When standards are used as part of rulemaking or implementing policy, the benefits and costs of the proposed rule must be assessed. OMB Circular A-4, Regulatory Analysis, provides guidance to government agencies on how to assess the costs and benefits as directed under OMB Circular A-119.

A-4 states that a "particularly demanding burden of proof is required to demonstrate the need for…mandatory uniform quality standards for goods or services if the potential problem can be adequately dealt with through voluntary standards…"

[17] Executive Order 13563, Executive Order 12886, and OMB Circular A-4. The Presidential Memorandum of January 20, 2021 directs the OMB to improve and modernize regulatory review. Changes to the review process are being considered as of the writing of this paper and may affect these EOs and the OMB Circular.

The Circular also discusses measuring the impact of regulatory action, including "Benefit-cost analysis (BCA) is a primary tool used for regulatory analysis."

Additionally, the Circular specifies that BCA and cost-effectiveness analysis (CEA) are tools that should be developed to analyze potential rulemaking when the primary benefits are improved public health and safety.

In essence, these U.S. Government requirements promote or require the assessment of the health and safety impacts of voluntary standards as part of an agency's rulemaking process.

3.4 Advocacy Organizations (Consumer, Labor, Environmental, Among Others)

Generally, a goal motivating participation by an advocacy organization is to promote the interests of its membership. These interests may include having the safest practical [product] available to their constituents, and to remove [products] that are unsafe or do not provide adequate safety from the market. The measurement of impact assists the organization to improve the overall safety associated with [products] in use and on the market.

As an example, a specific objective articulated by one consumer organization is to have a quantitative method that will enable them to understand when voluntary standards are sufficient and when standards need to be made mandatory.

Labor Unions promote the interests of their membership, including their health and safety. As an example, one union advocates for improvements in worker protections and safety regulation that may be based on voluntary consensus standards. By measuring the impact of standards, labor unions can assist in advocating for impactful standards to be adopted by management and government.

Environmental advocacy organizations are motivated to avoid adverse consequences that [products] may have on the environment. Impact measurement can help to demonstrate if standards are helping to achieve this goal.

4. Frameworks for Planning and Executing an Impact Assessment

Organizing an assessment of the impact of VCS is a complex undertaking. The intent of this section is to introduce and discuss several frameworks that may be useful in planning and executing an impact assessment effort.

4.1 Applicability of Different Types of Standards to Impact Assessment

There are a variety of [product] standards that impact human health and safety. Common content of standards may include elements such as ownership, scope, definitions, and revision process. Some standards may include example application and model adoption language. The following is not an exhaustive list but highlights the types of standards most relevant to health and safety.

- Product specification standards include requirements for end-products, components, and system standards.
- Codes set standards for constructed and installed objects, such as buildings and structures.
- Process standards address management, operations, services, and quality requirements.
- Test methods provide standardization of protocols that assist with the usefulness and comparative value of analytical methods.
- Terminology standards set forward consensus definitions, names, and verbiage to aid in clear communications for stakeholders.
- Personnel standards set qualifications to perform work or requirements to guide and evaluate work.
- Guides may be documents that provide standardized approaches to methods and processes.

At first glance, it may seem that the approach to measuring the impact these different types of standards would be very different. However, there are many common elements to an assessment. These elements will be explored in this section.

4.2 Framework: The Voluntary Standards Adoption Model

Assessing the impact of a standard requires an understanding of the lifecycle of the standard and a model describing the adoption of voluntary standards.

4.2.1 A Simplified Standards Lifecycle

The simplified lifecycle of a voluntary consensus standard includes development, revision, and retirement. This lifecycle is generally long, ranging from years to decades for different standards.

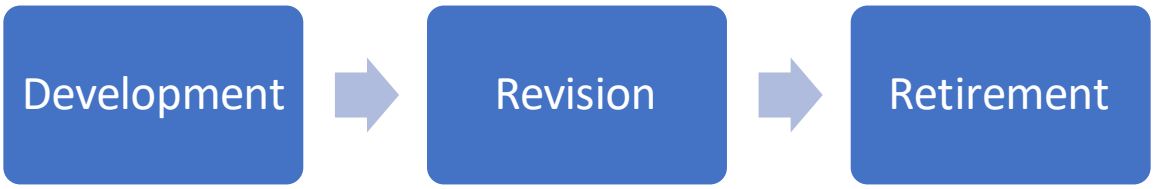

Figure 2 - Simplified lifecycle of a standard

Development: The development of a standard begins with the identification of the need and ends with the publication of the document by the standard owner. During this phase stakeholders are recruited to contribute to the effort; the scope, context, requirements, and other content of the standard are created and approved. The definition of the health and safety goals of the standard during this phase of the process will assist in the assessment of impact during subsequent lifecycle phases.

Revision: Typically, the requirements of a standard are changed during its life due to changing scope, characteristics, usage or applicability, environment, or the need for improved effectiveness. When revised, the health and safety benefits of the standard should also be reviewed for possible change.

Retirement: At some point, a standard may become obsolete or no longer necessary due to technological advancement, changes in society, or other factors. Standards may be retired or superseded; at this time, the standard is withdrawn from distribution.

4.2.2 A Voluntary Standards Adoption Model

To impact human health and safety, voluntary standards must be adopted and used by the target audience. The adoption of a standard is critically important to the impact that the standard will have. Because adoption has such a strong effect on impact, an expanded model of the voluntary standards lifecycle is required.

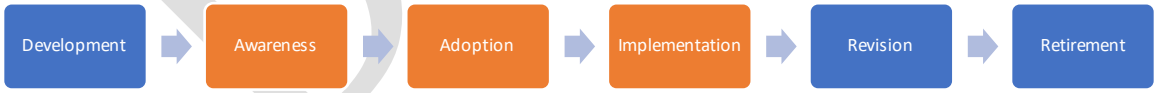

Figure 3 - Voluntary standards adoption model

The additional lifecycle stages include:

Awareness: Potential users must become aware of the standard if they are to begin the path to adoption and use. Creating awareness of standards in new, emerging, or modified technologies may be

particularly challenging. Impacted populations need to be aware of the existence of standards affecting their domains, not in detail, but enough to know they exist and to recognize implications of compliance.

Adoption: Adoption is the decision to apply the standard by the organization or regulator to meet its needs. For voluntary standards, this decision may be driven by a number of factors, including risk or liability mitigation, competitive pressures, or external drivers such as retailers or customers.

Implementation: The decision to adopt the standard leads to the stage of implementation. This is the process by which the requirements of the standard are applied to the [products] within the scope of the standard. Implementation of the standard implies that [products] begin to conform with the requirements.

4.3 Impact Varies During the Lifecycle of a Standard

During this lifecycle, a standard will have varying degrees of impact. The adoption of a standard by user groups, organizations, and national bodies will increase the potential impact of the standard.

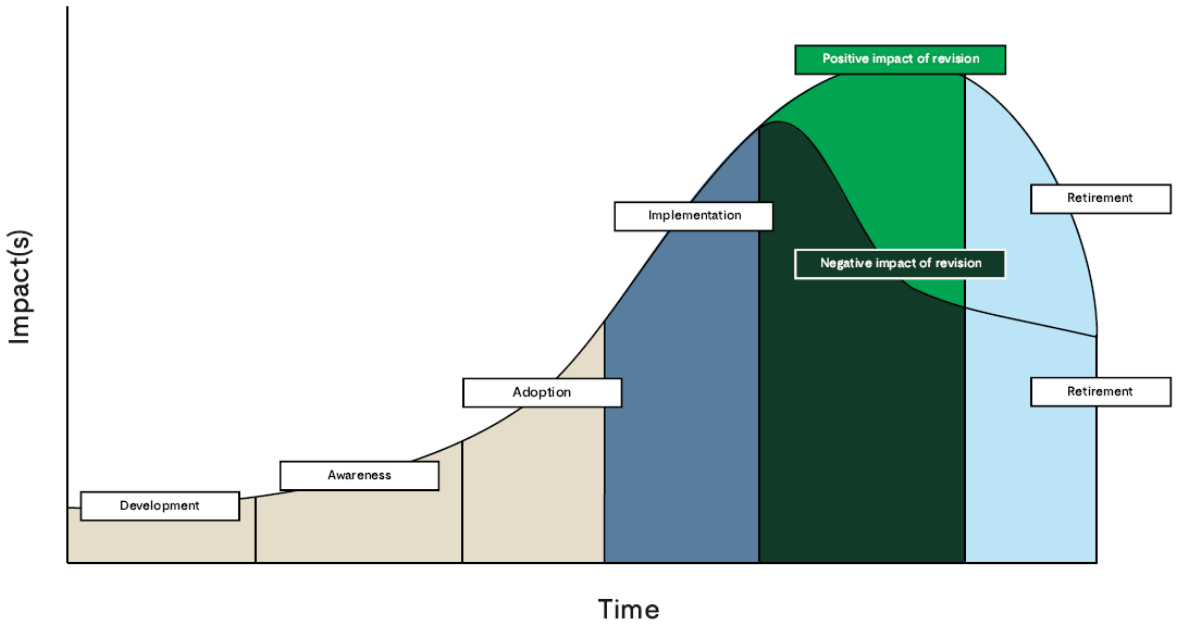

Figure 4 - Impact during the lifecycle of a standard

Development: A standard may have impact on human health and safety as the people and organizations that have discerned the need for the standard begin to implement the concepts and emerging requirements of the standard. Early adopters of safety concepts may also join in the use of these requirements while they are in development (i.e., Technical Committee member organizations). These impacts will be small, as only a few organizations will use the draft standard.

Awareness: Once a standard has been developed, organizations that will use the standard must become aware of the document. SDOs and industry organizations will promote the publication of the new standard. As more organizations become aware of the standard, they may begin to embrace the concepts of the standard as it may improve safety in these organizations. The impact of the standard remains relatively low, as the number of organizations aware of the standard, while growing, is low. Furthermore, because the standard has not been adopted, it can only influence the safety situation indirectly.

Adoption: This phase recognizes the decision of organizations, individually and collectively, to use the standard. Decisions to adopt a standard may be based on many factors, including regulatory requirements, market forces, and industry dynamics. As more organizations adopt the standard, impact grows. Even in the organizations that adopt a standard but have not yet implemented it, the decision to adopt begins the heart of the change cycle that results in impact.

Implementation: In this phase, organizations formally incorporate the requirements of the standard into the design, manufacturing, operation, and use of the [products] covered by the standard, or into their maintenance, contracting, documentation, hiring, training, and evaluation practices. As [products] and personnel conform to the requirements of the standard, its impact grows rapidly. Assuming the requirements of the standard are effective, the growth of impact increases with each [product] entering the market and each process implemented by end-users.

Revision: Almost all standards are revised during their lifecycle. These revisions are intended to improve the standard by addressing new developments in the market or to correct deficiencies in the existing standard. These changes to the standards, however, may have unintended consequences, and so the net change in impact may be positive or negative from the revision. It must also be noted that the revision to the standard goes through its own cycle of development, awareness, adoption, and implementation.

Retirement: Standards may be retired when they are superseded by new standards and/or as a result of the obsolescence of [products] or technologies. The impact of a standard may, in fact, outlive the actual standard, as [products] conforming to the standard may remain in use after the standard is retired and personnel and practices persist as they were until retrained. While this is the case for a limited time, ultimately the standard's impact returns to zero when the [products] conforming to the standard are retired. When personnel retire, though, they may be succeeded by people they trained; retiring and replacing personnel standards can follow a different trajectory than that seen in retiring and replacing [product] standards.

The process of assessing the impact of a standard should be tailored to the lifecycle stage of the standard in question. For example, a standard in the early phases of adoption will have a relatively small impact that may not be discernable. Leading indicators that indicate the potential for impact may be more appropriate measures at these early phases. A standard that has been adopted over time by many organizations will have more impact. At that stage, lagging indicators or specific outcome metrics will provide a more useful set of measures to assess impact.

4.4 Framework: Conducting an Impact Assessment

Figure 5 - An impact assessment framework

Preparation

Preparing to conduct an impact assessment includes the establishment of a charter for the assessment identifying the sponsor and audience(s) for the assessment, documentation of the health and safety objectives of the standard, and the preliminary identification of stakeholders.

Stakeholder Engagement

Best practices in assessing the impact of programs stress the identification and engagement of stakeholders in the process. Strong stakeholder engagement can help with robust planning, measurement, and analysis of the assessment. Stakeholders provide important points of view on the theory of change, timing, metrics selection, and data sources. Furthermore, by involving stakeholders in the assessment process, the results of the assessment will likely have greater buy-in from those stakeholders.

Planning

Development of an impact measurement plan is critical to success. While the plan must be flexible in order to incorporate changing circumstances and lessons learned, measurement is most appropriately based on understanding the intended outcome of the standard or its change. This hypothesis for impact should guide the remainder of the assessment process.

Scoping: Understanding the scope of the standard and how it relates to the standard's place in society is critical. There are several dimensions related to the scope that need to be evaluated. First, what is the technical scope of the standard (e.g., electrical safety, professional qualification, etc.)? What is the scope of its applicability? For example, does a product safety specification standard apply to all products in the market, or to a specific subset (e.g., medical devices)? Does it apply only to products used in specific environments (e.g., medical devices intended for home use)? Considering these questions is important in order to answer the question "Where should I look for impact?"

Rationale: When new standards are developed, or changes to standards are proposed, they are put forward with a basis or rationale to justify their development. This rationale may be formal or informal,

but it will provide a foundation for the "theory of change."[18] Additionally, a review of the rationale will help to delineate the hazards to be studied during the assessment. Many standards address multiple hazards, and the plan must clearly specify which hazards (all or a subset) are to be evaluated. This answers the question "How will the standard create impact on matters under its scope?"

Timing: VCS normally have a long lifecycle. Understanding the timeline of the standard or a change and the affected [product] will inform the analysis. In the early stages of the lifecycle, one must use leading indicators; at later stages lagging indicators will be present. This answers the questions "What kind of impact will the standard have?" and "When will there be impact?"

[Product] Lifecycle Considerations: The [product] will also have its own lifecycle that will affect timing. For instance, short-lived, consumable products in the market will be replaced more quickly than durable products. Therefore, standards affecting durable products will take longer to achieve the desired impact.

Metrics Selection: From the rationale, one can derive potential metrics to measure the impact. These metrics may measure the progress towards impact—for example, measuring the level of awareness, adoption, or use/enforcement of a standard—or they may directly measure the desired outcome of implementing the standard. An example of the direct measure may be rates of occupational injury related to falls from height.

Measurement

Baseline measurement: In order to assess the impact of a standard, one needs to understand the baseline situation that was in place before the standard came into effect. The baseline measurement approach will be informed by the rationale and metrics selection. Metrics selected for the baselining phase will be used again for the evaluation. This answers the question "What was it like before the standard?"

Target State Measurement (Data Collection): The assessment of impact will be derived from the change in metrics from the baseline. This requires that the same metrics are used in the measurement phase as those in the baseline phase. Based on the timing of the standard and the [products] covered, there may be multiple measurements taken over time to fully assess the impact of the standard. This answers the question "What is it like now that the standard is in place?"

Analysis

The analysis of the situation before, during, and after the promulgation of a standard provides the core of the measurement of impact. The measured change between the baseline and "current" state determines whether impact has been achieved. However, this analysis may not be straightforward.

[18] "Theory of Change is essentially a comprehensive description and illustration of how and why a desired change is expected to happen in a particular context." Center for Theory of Change, accessed May 7, 2023 https://www.theoryofchange.org/what-is-theory-of-change/

Confounding factors, including market forces and other mitigating actions, may hide or amplify the impact of a change, and should be evaluated during the analysis.

Reporting

A report should be prepared to document the assessment process, scope, methods, summary data, analysis, and conclusions. The report should stand on its own, providing the audience with sufficient detail to allow them to understand how the conclusions regarding impact were reached, and the degree of confidence in those conclusions. The report may also serve as a reference for follow-on impact assessment work.

5. Impact Components

Impact may be thought of as a function of the two factors of effectiveness and conformance.

$$Impact = Effectiveness \times Conformance$$

Effectiveness and conformance are the methods of achieving an impact. While it may be possible to measure each individually, it is better to treat each as a leading indicator for impact. Effective standards do not create impact without broad conformance to the standards. Improving conformance to an ineffective standard also does not create impact.

5.1 Examining the Effectiveness of a Standard

Effectiveness represents how successfully the requirements in the standard, as written, would address or mitigate threats to human health and safety such as illness and injury. For example, a more precise test method for contaminates in the air may produce improved health and safety compared to a less precise test method.

In the context of human health and safety, the most effective standard would be one that, when complied with, eliminates the hazard. Those standards that do not eliminate the hazard but provide robust mitigations to the hazard (e.g., physical barriers, reduction in emissions, improved detection, etc.) may be highly effective as well. Less effective measures may include requirements to warn to users of hazards, administrative or procedural controls, or training. When evaluating the effectiveness of a standard, one should review the overall system of hazard reduction strategies employed. Multiple, defense-in-depth strategies (e.g., physical barrier + administrative controls + training) are more effective than individual strategies (e.g., administrative controls alone).

The effectiveness of a standard/revision comes through the development and maintenance of the requirements of the standard. As written requirements are removed, added, or changed in the standard, the effectiveness of the standard changes. The standards committees that oversee the drafting of the standard and the other stakeholders that submit comments and proposed changes help keep the standard as effective as possible. In general, standards committees should mirror the diverse range of stakeholders impacted by the standard. [19]

5.2 Examining the Conformance to a Standard

Conformance is directly related to the overall number of conforming [products] the impacted population has in its possession. Conformance is a composite factor characterizing both (a) the degree to which the standard is adopted or how broadly is the standard is used; and (b) the degree of adherence to the

[19] See, for example *ANSI Essential Requirements*, *Guidance on Balance and Outreach within the American National Standards (ANS) Process* (ExSC_013_2022) and *Guidance on Lack of Dominance within the American National Standards (ANS) Process* (ExSC_013_2021).

standard or how precisely the standard is followed. Means to achieve improved conformance to a standard include:

- changing the motivation to conform through requirement by law or code
- by incorporation by reference into another standard
- by industry mandate such as a retailer or supplier
- by the manufacturer's desire
- or through the assurance of conformity, through improved verification, frequency, inspection, tools, or training of personnel.

Additionally, the time needed for the end users of the [product] to replace existing, non-conforming [products] with new, conforming [products] should be accounted for. Conformity assessment is a tool that can be used to assess the overall conformance that [products] have to a standard. However, if the conformity assessment status of all [products] is not known, one should analyze underlying factors of conformance to understand its role in measuring the impact.

5.3 Impacted Population

The impacted population is the group of people whose health and safety are directly or indirectly affected by the hazards the standard mitigates. For example, the impacted population of a consumer product could include the end user consumers, those that make, sell, install, service, and dispose of the product, and those affected indirectly such as family, community, and caretakers. The impacted population of an environmental test method may be the entire population of a community where the test method is used to monitor the water quality of a municipal supply.

5.4 Approaches to Measurement

Assessing the impact of a standard implies a quantitative approach, yet this may lead to an overreliance on numeric data, which may not be available at the time the assessment is performed, or in the quantity and of the quality necessary for valid calculations.

Quantitative research: Where possible, it is preferable to measure the impact of a standard directly through quantitative methods. These methods will focus on the direct measurement of the desired effect of the standard (e.g., failure rate, hazard mitigation, etc.) and the conformance of [products] to the standard (e.g., through market data).

Quantitative measures may include incident counts, incident rates, questionnaire and survey responses, product surveillance results, sales/distribution of standards, and standards training enrollment.

Qualitative methods: There are many situations where the direct measurement of the effectiveness or conformance of the standard are impractical. Qualitative methods include the gathering and evaluation of non-numerical data. The assessment framework enables the use of qualitative measures to estimate the impact of the standard. Some sources of qualitative data include use of: subject matter experts'

judgment; interviews and focus group data on attitudes and perceptions of risk or performance; document reviews of related matter (e.g., news articles); and, observation of personnel performance of duties or processes. Note that these may yield numeric results though they do not directly represent safety outcomes.

Mixed-methods research: Mixed methods systematically integrate quantitative and qualitative approaches to research in order to answer research questions. Mixed-methods research is pragmatic, collecting both narrative and numerical data, employing both structured and emergent designs. It , analyzes data both via statistical and content analysis, and makes meta-inferences about answers to research questions by integrating the inferences gleaned from qualitative and quantitative findings. [20]

[20] A. Tashakkori, I. Newman, in International Encyclopedia of Education (Third Edition), 2010.

6. Guidance on Measuring Impact

Measuring the impact of a standard is a highly specific task. Each [product], industry, hazard, impacted population, and circumstance may need specific information that cannot be summarized in a single document. Therefore, this paper provides generalized guidance on how to conduct an impact assessment. The most important information needed to measure impact is data. Data on injuries, disabilities, and illnesses within the impacted population are the obvious choice of data to directly assess the impact of standards on public health and safety, but other metrics may be worth considering as well. Data on adverse incidents that did not result in injury but could have under different circumstances (i.e., near misses), cases of exposure to known pathogens without manifestation of disease, measured reduction of substances known to—directly or indirectly—impact health and safety, and healthcare costs (as a proxy for data on prevalence of injuries and illness) are some examples of alternative potential data for measuring impact. Using the data, one can create a baseline, compare the changes in the data to the baseline, consider the form of analysis, and, finally, consider the confounding factors.

6.1 Selecting or Collecting Data

To start, one must first know where the data for the outcome measures will come from. Will you collect it for the assessment, or does it already exist? Data on a population is typically collected through either a survey or a census process. A survey collects data from a sample of the whole population and then attributes the results to the entire population. The most important question to ask when considering sample data (or preparing to take a sample) is how well does the sample represent the impacted population. A census collects data from the entire impacted population. Census data is much less available due to the expense and difficulty in conducting it. The most important question to ask when considering census data that may be available (or preparing to conduct a census) is its rigor. Are the data collection mechanisms sufficient to credibly claim that data was collected from the entire population within an acceptable margin of error?

Some factors to consider when considering how well a sample represents the impacted population include:

- Is it large enough?
- Do the demographic characteristics of the sample look like the impacted population as a whole? This could include gender, age, racial, ethnic, physical, mental, cognitive, or developmental differences in the population as a whole.
- Do the social determinants of health of the sample look like the impacted population as a whole? This could include economic stability; education access and quality; healthcare access

and quality; neighborhoods and built environments (e.g., safe neighborhoods, transportation, etc.); and social and community context (e.g., relationships with friends, family, etc.)[21].

Accessing representative sample data can often be a challenge. Frequently, assessments must rely on samples of convenience, or those that are not representative of the population due to resource or other constraints. This will limit the conclusions that can be drawn from the data. In these cases, it is crucial to explicitly acknowledge the limitations of the data, discuss the implications of those limitations, and clarify if caution needs to be exercised when extrapolating results to the population in question or other populations. As an example, available credible sample data that may be useful for measuring the impact of consumer product safety standards include the BLS Survey of Occupational Injuries and Illnesses, and NEISS.

Census data come from an actual count of the entire population, or the entire impacted population. Representativeness is not an issue with a census, but they are relatively rare and difficult and costly to conduct. A challenge is ensuring that the entire population or impacted population is counted. As an example, existing census data that may be useful to measure the effectiveness of consumer product safety standards include the BLS Census of Fatal Occupational Injuries (CFOI) and the CPSC Death Certificate (DTHS) database (part of the CPSC Clearinghouse).

6.2 Establishing a Baseline

With the appropriate data, one must next determine an appropriate chronologic baseline to compare that data against. The baseline should occur before the introduction of the standard (or a change to the standard). This may be before the publication of the standard if conformance to the standard begins before publication of the standard. For instance, employers may choose to implement new worker safety requirements before the associated standards incorporate those requirements.

Some factors to consider when developing a baseline include

- Does the baseline cover a sufficient timeframe?
- Do other relevant factors cause the baseline to change over time?

6.3 Determining the Outcome

The outcome is defined as changes in the data after publication of the standard, when one believes there is sufficient conformity to the standard or revision (with sufficient diffusion) to produce a change. Sufficient diffusion of the standard should be determined based on the factors appropriate to the [product], industry, standard, hazard, and impacted population. The outcome time intervals should be at least comparable to, if not, longer than the baseline time scale. The time scale should be sufficient to accommodate any lag in the time it takes for the diffusion of the replacement [products] into the

[21] U.S. Department of Health and Human Services, Social Determinants of Health, accessed May 7, 2023 https://health.gov/healthypeople/priority-areas/social-determinants-health.

population. This time lag for diffusion will vary depending on many factors with the estimated lag time increasing proportionally to expected life and replacement cost.

6.4 Form of Analysis

One consideration in measuring outcome concerns the form of the data. Raw counts, or more precisely changes in raw counts, are a weaker measure of impact than changes in proportions, ratios, or percentages. To demonstrate the potential weakness with the use of changes in incident counts as an outcome measure, imagine a hypothetical consumer product~~,~~ Product X, was observed to experience no decrease in the number of incidents of harm over a one-year period following the publication of a revised standard and appropriate diffusion period. It may first appear that the revised standard is not effective in improving health and safety. However, if at the same time Product X had a great sales year and doubled the number of units in the market, then the actual injury-to-product ratio would have dropped by half. This represents a 50% reduction in the occurrence of injuries per product. Most would consider this outcome a substantial improvement in health and safety.

Incidence rate, a common outcome metric used in injury epidemiology, represents the number of incidents divided by the impacted population. This has traditionally counted occurrences of a specific injury type but the incidence rate could be calculated for any type of incident including injuries of a specific severity (e.g., hospitalizations), near misses, or even incident costs (e.g., medical costs associated with specific incidents). However, calculating the impacted population of most health and safety standards can be very difficult. Even a credible estimate is beyond impractical in most situations. Two alternatives to incidence rate that may be useful are: 1) taking a representative sample and calculating the sample incidence rate (and imputing that rate to the entire population), and, alternately, 2) using a proxy metric for incidence rate.

One possible proxy measure is the number of incidents per the number of [products]. This would be similar to the concept of defect rate in quality control.

6.5 Confounding Factors

A change in the outcome metric (e.g., decrease in incidents) from before to after the introduction of a standard or revision is necessary, but not sufficient, to demonstrate the standard is effective. Other, confounding factors may influence the change. Measuring the impact by analyzing the data before and after a change to a standard does nothing to control all the other potential sources of the change. For instance, did other [products] in use significantly change how the [product] under question is used? Did characteristics of the end users of the [product] change how the [product] under question is used over time? In scientific experimentation, all the alternative potential causes for a change to the dependent variable (i.e., the outcome metric) are controlled, but this is not possible with an impact assessment. Instead, means to discount alternate explanations are needed. Some examples of alternate methods include: 1) repudiation of potential contrary explanations through examination, 2) data triangulation using multiple sources of independent data yielding confirming results, and 3) use of a logic model that

describes the process by which compliance with the requirements of the standard will enhance health or safety that is supported by evidence.

7. Conclusions

Many stakeholders are convinced that voluntary consensus standards have significant positive impacts on the health and safety of the public. Anecdotally, there are many case studies and reports describing the impact that specific standards (or their revisions) have had on hazards within the scopes of these standards. Moreover, the organizations developing standards and the many subject matter experts who contribute to the process have a general impression of the impact of their standards through feedback received from the users of the standards, e.g., manufacturers or consumer advocacy groups. However, quantifying the impact of voluntary consensus standards proves to be surprisingly difficult.

Voluntary consensus standards for health and safety are diverse in their scope, technical approach, industry, and user base. As the Working Group discussed the variation between these types of standards, it became clear that this imposes limits on the ability to specify one single approach to measuring the impact of standards on health and safety. This necessitated that the group focus on a set of concepts, frameworks and considerations that may be used to develop a specific approach based on the standard being assessed and other variables such as the availability of data. This white paper provides a starting point for an impact assessment.

Some of the key concepts, frameworks and considerations include:

Impact can be summarized in the combination of two critical elements: conformance and effectiveness

$$Impact = Effectiveness \times Conformance$$

Effectiveness represents how successfully the requirements in the standard, as written, would address or mitigate threats to human health and safety such as illness and injury. Conformance is a composite factor characterizing both (a) the degree to which the standard is adopted or how broadly is the standard is used; and (b) the degree of adherence to the standard or how precisely the standard is followed.

The impact of a standard changes over time. The voluntary standards adoption model means that conformance occurs over time, and that time is related to the diffusion of the standard across the industries or sectors it affects. Furthermore, the products, systems, components, processes, and people subject to the scope of the standard may pre-date the requirements and thus reduce overall conformance.

Assessing the impact of a standard requires a structured framework. Critical elements of a structured framework include stakeholder engagement, planning, measurement, and analysis.

Data gathering or measurement can be particularly challenging in the impact assessment process. Direct measurement of injury or illness associated with the specific requirements of a standard is very difficult due to the limitations of existing data sources. Furthermore, the use of incident rates is preferred to incident counts, but data for normalization can also be difficult to obtain. Finally, confounding factors

must be addressed to establish the relationship between the change in baseline with the measured outcome, otherwise false conclusions may be made. Because direct outcome data are difficult to obtain, proxy data such as economic costs of injury may be useful to estimate the impact of a standard. Additionally, leading indicators of impact may be helpful in the qualitative assessment of impact. Leading indicators may include data on adoption of the standard.

The benefits of measuring the impact of voluntary consensus standards on human health and safety are significant. These benefits include the alignment of resources and investments to those efforts that have the highest impact, improved management of a portfolio of standards, and communicating the overall benefits of standards development activities to stakeholders.

Appendix A. Definitions

Conformance: a composite factor characterizing both the quantity of [products] meeting the standard, and the degree of the adherence of these [products] to the standard.

Conformity assessment: techniques and activities that ensure a [product], installation, project, data, design, material, claim, person, body or organization or, or any combination thereof, fulfils specified requirements.[22]

Consensus: Consensus means substantial agreement has been reached by directly and materially interested parties. This signifies the concurrence of more than a simple majority, but not necessarily unanimity. Consensus requires that all views and objections be considered, and that an effort be made toward their resolution.[23]

Consensus body: The group that approves the content of a standard and whose vote demonstrates evidence of consensus.[24]

Effectiveness: The extent to which planned activities are realized and planned results are achieved.[25] For the purposes of this paper, effectiveness is the ability of the standard to produce a desired change in the impacted population.

End User [product]: The person who uses the [product] that is the subject of the standard. This may or may not be a member of the impacted population.

Impact: Actual change(s) (both positive and negative) that the standard has produced on the impacted population. In the context of this paper, these changes may generally be categorized as mortality, morbidity, or property loss.

Impacted population: The impacted population is defined as the group of people whose health and safety are directly or indirectly affected by the hazards the standard mitigates. It is critical to note that the **impacted population may be the end user [product] but this will not always be the case.**

[Product]: A [product] is the output of an organization that can be produced without any transaction taking place between the organization and the customer or end user [product].[26] **This includes physical products, processes, services, methods, personnel, and other intangible assets that are the subject of a standard.**

[22] What is conformity assessment?, ISO, Conformity Assessment (iso.org), accessed May 5, 2023.
[23] *ANSI Essential Requirements*, January 2022.
[24] *ANSI Essential Requirements*, January 2022.
[25] ISO 9000:2015, Quality management systems — Fundamentals and vocabulary, 2015.
[26] ISO 9000:2015, Quality management systems — Fundamentals and vocabulary, 2015.

Standard owner: The organization that is responsible for the development and maintenance of the standard.