The BC Patient Centred Measurement Working Group

Emergency Department 2018

Healthideas Toolkit for Data Users

Version: 1.0
Created: June 2019

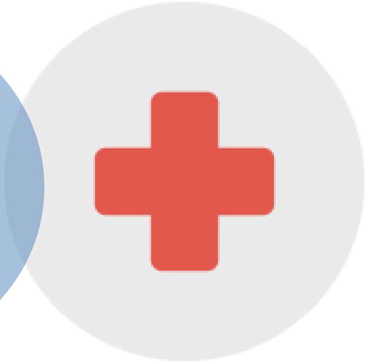


Table of Contents

The BC Patient-Centred Measurement Working Group thanks those British Columbians who participated in this survey, providing valuable information about their lived experiences to support clinicians and policy makers in their efforts to provide the highest quality of care possible for all patients cared for in BC hospitals.

About the Toolkit: Emergency Department Survey 2018

This document contains information and supporting materials in order to provide users of the Emergency Department 2018 patient-reported experience and outcome survey results with sufficient context to make informed use of the data provided through Health*ideas*. This document does not replace any technical documents; rather, it serves as a complementary source of information.

The document will be revised as necessary should additional information and materials become available.

For details about the survey, Health*ideas*, or if you have any additional questions, please contact:

Lena Cuthbertson
Provincial Executive Director
BC Office of Patient-Centred Measurement & Improvement

✉ lcuthbertson@providencehealth.bc.ca

☎ Office: 604-806-9401

About the BC Patient-Centred Measurement Working Group

Since 2003, the BC Patient-Centred Measurement Working Group (BCPCMWG), which includes representation from the BC Ministry of Health, all seven BC Health Authorities, and their affiliate organizations, have implemented a program to measure the self-reported experience, satisfaction and health-related quality of life of the people who use a range of healthcare services in BC.

These surveys are provincially coordinated and conducted across all locations of service, i.e., province-wide. The survey instruments and the results are based on a scientifically rigorous process for learning from patients and improving the quality of the healthcare and services provided in BC. Clinicians, leaders, policy makers, and, most importantly, patients are involved at every step of the planning of these surveys, including the development and testing of questions, the selection of survey instruments, and the validation of results.

> "This initiative is about giving people who use British Columbia's (BC's) health services the opportunity to provide feedback about their experience and satisfaction with the care and services they receive, as well as providing information about their outcomes and health-related quality of life."

The results of surveys that ask users of the health care system in BC for feedback are intended to be used by Health Authority clinicians and leaders to improve the quality of the experience and the clinical outcomes of the patients, residents, and families at the point of care and to promote continuous organizational improvement. Additionally, Ministry of Health and Health Authority executives and policy makers are interested in survey results as an accountability measure to understand the performance of the health care system at individual and cross regional and provincial levels.

To date, province-wide surveys have been conducted in BC in the:

- Emergency Department sector (2003, 2007, 2009 to 2015, 2018)
- Long-term Residential care sector (2003/04 and 2016/17)
- Acute Inpatient sector (maternity, pediatrics, surgery (2005, 2008, all this plus rehabilitation in 2011/12, and again in 2016/17)
- Short stay Mental Health (inpatient psychiatric units) and Substance Use (detox, support recovery, and withdrawal management) sector (2010/11)
- Outpatient Cancer Care sector (2005/06, 2012/13); Cancer Transition to Survivorship (2016)

About Health*ideas*

In 2011, the Health Information Privacy and Security Council (HIPSC) agreed to the practice of returning raw survey data, including patient identifiers, back to each health authority or to the affiliate organization where the patient received care for purposes of secondary data analysis and to inform quality improvement. What was missing was the ability to link the survey data to other clinical and administrative data sets and to analyze the data beyond a single health authority (i.e., at a provincial level or across health authorities).

In July 2014, the HIPSC agreed that the BC Ministry of Health's Health*ideas* data warehouse could be used to centrally store patient self-reported survey data. Health*ideas* is a safe and secure source of information that was created and is managed by the BC Ministry of Health. Health*ideas* was designed to support decision making and contains information about hospital and physician services, population data, and other reference data.

Health*ideas* will act as the repository of all survey data collected from BC patients, clients, residents, and families. It will host all the records of patients with an encounter in any of the sectors surveyed, flagging those who were sampled and invited to participate in a survey, as well as all those who completed a survey. Each approved user will be provided with a specific level of access, based on need and authority.

Purpose of the 2018 Emergency Department Survey

This survey asked patients about their health-related quality of life and their experiences with the quality of the care and services received as a patient in one of 108 Emergency Departments in BC. Patients who visited an Emergency Department between January 1, 2018 and March 31, 2018 were eligible to receive a survey.

As with all provincially coordinated surveys conducted by the BC Office of Patient-Centred Measurement on behalf of the BCPCMWG, the Ministry of Health and health authorities are committed to use the survey results to:

- Enhance the performance of the Emergency Department sector in the province;
- Enhance public accountability;
- Support quality improvement initiatives; and
- Contribute information to support research and researchers.

Understanding patients' experiences and self-reported health related quality of life is vital in ensuring BC's health care system is meeting the needs of patients – allowing them to become partners in their own health care.

The aggregate results of the Emergency Department 2018 survey were disseminated to staff and leaders across the BC health authorities in May 2019.

The goal is that the results that reflect the "voices" of BC's patients will be used improve the experience and out care for all patients in BC.

Glossary of Terms

Dimensions or Domains of Patient Experience

From a psychometric point of view, the terms dimensions and domains of patient experience are interchangeable terms used to describe a group of items that are being evaluated in a survey. A summary score is often calculated to quantify the dimension, with the score being a composite score of the questions that make up of that dimension. Dimensions and domains of patient experience may be conceptually derived (individual items make intuitive sense to be grouped) or empirically derived (individual items have been shown to fit together statistically).

Factor

In the field of measurement and psychometrics, a factor is an indirect representation of the underlying dimension inferred from the question (item) responses. Mathematically, a factor is a weighted linear combination of items (e.g., survey questions) thought to summarize the variations observed in the item responses. Also known as 'latent variable' in statistics, a factor is considered to be unobservable but is inferred from items that are considered observable (i.e., directly measured).

Freedom of Information and Protection of Privacy Act

The BC Freedom of Information and Protection of Privacy Act (FOIPPA) protects the personal privacy of BC citizens by prohibiting the unauthorized collection, use, or disclosure of personal information by public bodies.

Item and Question

The words 'item' and 'question' are often use interchangeably; however, the term 'item' is more broadly defined as not all items are phrased as questions. In the field of measurement or psychometrics, survey questions are referred to as items. Items can refer to things such as multiple choices, statements, ratings assigned by an observer, and performance assessment. An item bank is a collection or repository of items.

Key Driver

A key driver is a survey question that reflects aspects of care and service shown to statistically have the greatest influence on the global rating indicator questions. Ratings of overall experience and likelihood to recommend are examples of global ratings.

Margin of Error

The margin of error is an indicator of survey accuracy that measures the imprecision inherent in survey data. Margin of error is inversely related to the sample size used to draw inferences about the larger population. A margin of error of ± 5% is considered good while ± 8% is acceptable.

Multiple Imputation

Multiple imputation is a statistical method that is widely recommended as a robust approach for minimizing biases due to missing values in statistical analysis. It creates multiple copies of data with imputed values and pools the analyzed results into a single estimate.

Norm-Referenced Score

A score is norm-referenced if it is interpreted with regard to the performance of a peer group, a reference population, or benchmark. Percentile rank is an example of a norm-referenced score.

Psychometrics

Psychometrics is a scientific field of study concerns with the theory, practice, and techniques of psychological and behavioural measurement. This includes improving the measurement of knowledge, abilities, attitudes, opinions, and personality traits via the development of assessment tools, statistical methods, and mathematical techniques.

Patient-Reported Experience Measures

Patient-reported experience measures (PREMs) are measurement instruments that patients complete to self-report their global ratings of overall satisfaction with the care and services received and their experiences with the processes of their care.

Patients-Reported Outcomes Measures

Patient-reported outcome measures (PROMs) are measurement instruments that patients complete to self-report information on aspects of their health status that are relevant to their quality of life, including symptoms, functional, physical, mental and social health.

Privacy Impact Assessment

A Privacy Impact Assessment (PIA) is a process used to evaluate potential impacts of a study or program on participants' privacy rights and to ensure compliance with privacy protection rules and responsibilities. Completing a PIA is a legislative requirement when developing or changing a system, project, program, or activity. A PIA is conducted for all BC PCM Working Group initiatives, projects, and surveys; each PIA is reviewed by the Health Information Privacy and Security Operations Committee (HIPSOC), which is a sub-committee of the Information Privacy and Security Standing Committee at the Ministry of Health, and the Office of the Information and Privacy Commissioner.

Questionnaire

A questionnaire is a form of data collection, asking an individual to respond to a set of printed or written questions with a choice of answers to gather information from respondents. Questionnaires can be administered in-person, online, by phone or mail.

Reliability

Reliability is a measure of the repeatability or consistency of results obtained from a standardized survey instrument. A survey instrument itself is neither reliable nor unreliable; it is the responses that can be consistent or repeatable. In addition, just because a response to a scale is reliable does not mean that it is valid, that is, it measures what it is supposed to measure (see Validity).

Response Rate

Response rate is the number of people who answered ("completed") the survey divided by the number of people in the sample. It is usually expressed as a percentage and is one of the most commonly used indicators to gauge the quality and accuracy of survey data. There are different response rate calculation standards, with varying definitions of "answered survey" or "complete" and who to include in the sample.

Scale and Sub-Scale

In the field of measurement or psychometrics, a scale is a collection of items (e.g., survey questions) designed to measure one dimension. Similarly, a sub-scale is a subset of items from that same item collection use to measure a particular aspect or component of that dimension (sub-dimension). A standardized survey instrument is often designed with items that form a scale.

Statistical Significance

A statistically significant result (that represents a difference between two groups of scores) is a result that is unlikely to have occurred due to chance if there really was no difference between the two groups of scores. In other words, a statistically significant result occurs when the difference between two groups of scores is large enough that we can say that the probability of this difference occurring is very small if there really is no difference in scores between groups. A statistically significant result may or may not be relevant (i.e., "practically significant") in a clinical context.

Structural Equation Modeling (SEM)

Structural Equation Modeling (SEM) is a statistical technique that combines psychometric factor analysis and multiple regression analysis in a single statistical model.

Survey

A survey is a process for gathering information that could involve a wide variety of data collection methods, including a questionnaire. It could also involve observing or measuring things that go beyond questions, including physical measurements or judgments made by an observer. A survey typically includes questions from one or more questionnaires or instruments that address specific objectives and may also be used to collect demographic information.

Survey Instrument

A survey instrument is a tool that follows scientific protocols for obtaining information from respondents. For survey research, the survey instrument often involves a questionnaire that provides a script for presenting a standard set of questions and response choices.

Survey Vendor

To carry out the work of the BC Office of Patient Centred Measurement, contracts with external research companies are negotiated on behalf of the health authorities. Depending on the size of the contract, the provincial group will go through a procurement process facilitated by BC Clinical and Support Services (BCCSS). All survey vendors are required to adhere to strict privacy and information security requirements, as specified by applicable BC legislation.

Survey Weights

Survey weights are used to make the sample representative of the target population on key characteristics such as organization level attributes or demographic characteristics during analysis. Survey weights, or the inverse probabilities of selection for each observation, allow users to reconfigure the sample as if it was a simple random draw of patients that is representative of the total patient population to yield accurate estimates.

Top-Box Score

The top-box scores is the percentage of respondents who selected the most positive response category to a survey question (e.g., the 'Always' response option from the choices Never, Rarely, Sometimes, Usually, Always).

Total Valid

The Total Valid number is the number of respondents who reported a valid answer (i.e., excluding missing and not applicable responses) for the question.

Valid Percent

The percentage of responses based on the Total Valid number (i.e., excluding missing and not applicable responses).

Validity

Validity typically speaks to the accuracy of an assessment tool in terms of, whether or not it measures what it is supposed to measure. A survey instrument or item may be reliable but may not necessarily be a valid measure. Validity must be formally established by empirical studies as well as sound psychometric and test development practices. The definition of validity itself has been subjected to debate. In particular, the current Standards for Educational and Psychological Testing (developed jointly by the American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education) champions the view that a survey instrument is neither valid or invalid (i.e., validity is not a property of the test). Instead, validity is defined as the degree to which ongoing empirical evidence and theory support the conclusions drawn from the survey instrument for its intended purposes.

Variable Importance Analysis

Variable importance analysis is a class of methods that examine the relative importance of predictors in a statistical model. One such approach, the Pratt Index, quantifies the relative contribution of an item or dimension to the total explained variance of the outcome. In the ED report, the Pratt index was used to identify key drivers from the SEM analysis.

Selected Survey Tools

The *2018 Emergency Department Sector Survey* marks the first time that information from patient-reported experiences of care measures (PREMs) has been collected simultaneously with patient-reported outcome measures (PROMs) related to an emergency department visit. The below table summarizes each of the question blocks or modules included in the survey.

Survey Section	Question Block or Modules	No of Questions
PREMS	The Emergency Department Patient Experiences with Care (EDPEC): • EDPEC Discharged to Community Instrument (ED_DTC) • EDPEC (Admitted Stand Along Instrument (ED_ADMIT: EDPEC_IP1 and EDPEC_IP2)*	 35 2
	BC Emergency Department questions (BCED)	18
	"Hello my name is" questions (QABED)	2
	Emergency Health Services questions (BCEHS)	6
	Patients who saw a doctor identifier (DR_SCREEN)	1
	BC's Patient Safety Module • Hand hygiene question bank • Medication reconciliation question bank	 6 3
	Office of the Seniors Advocate questions (OSA)	5
	BC's Continuity Across Transitions in Care Module (CONT)	14
	Intravenous Vascular Access questions (IVT)	8
	BC Emergency Medicine Network questions (EMN)	16
PROMS	EQ-5D-5L**	6
	Veteran's Rand 12 Item Health Survey (VR-12)	14
DEMO	Demographic questions	2
APL	Aboriginal Patient Liaison questions***	2
OPEN	Open-ended patient comment question	1

*If the patient was admitted to acute care only.
** Age ≥ 13 years only.
*** Northern Health Authority patients only.

Emergency Department Patient Experience of CARE (EDPEC):

The EDPEC is a survey designed to understand patient experiences of emergency department care and was selected as the core instrument for seeking feedback from patients who made use of BC's Emergency Health Services System, including transport by ambulance. Two versions of EDPEC were included: The EDPEC Admitted Stand Alone Instrument (ED_ADMIT), which has questions specific to patients who were admitted to the hospital following their emergency department visit; and the EDPEC Discharged to Community Instrument (ED_DTC), which has questions for those who were discharged directly to the community.

The Veteran's Rand 12 Item Health Survey and EQ-5D-5L:

The VR-12 is a generic (i.e. not condition specific) patient reported outcomes measure (PROM) that focuses on self-ratings of health-related quality of life. It is an abridged version of the Veterans RAND 36-item Health Survey (VR-36). The VR-12 includes questions that will produce scores for the following:

- Overall health status
- Physical health status
- Mental health status
- 7 health domains including:
    - physical functioning, social functioning, energy-fatigue, bodily pain, role limitation, perceived general health, and perceived mental health

The EQ-5D-5L is a generic PROM developed by the EuroQol (EQ) Group that measures self-reported health-related quality of life. The original descriptive system measures five dimensions (5D) using five response levels (5L: no problem, slight problems, moderate problems, severe problems and extreme problems):

- Mobility,
- Self-care,
- Usual activities,
- Pain / discomfort,
- Anxiety/depression)

The EQ-5D-5L results were not publicly reported, as the data was used to inform decisions about a generic PROM instrument for use in BC and nationally in Canada. While other studies have examined the psychometric properties and validity evidence pertaining to the use of these PROM measures in various settings and populations, the inclusion of PROMS in the BC Emergency Department survey was motivated by the questions "What kinds of information do these PROMs provide? What 'stories' do they tell?". The overall goal was to help inform the selection a generic PROM for use in BC and in Canada to measure the self-reported health-related quality of life and health status of individuals who use healthcare services.

Sample Plan

Patients who received Emergency Health Services from one of BC's 108 emergency departments, urgent care centres, and diagnosis and treatment centres between January 1, 2018 and March 31, 2018 were eligible to participate in the 2018 Emergency Department Patient Survey.

While the intent of the 2018 Emergency Department Sector Survey was to capture as wide a range of patient experiences as possible, not all encounters were eligible for inclusion. Exclusion criteria were as follows:

- Patients deceased in hospital (either in the ED or following admission to an Acute Care unit);
- Infants less than or equal to 10 days old at discharge;
- Patients with no fixed address and no telephone number (patients with only a phone number were included);
- Patients residing outside of British Columbia; and
- Patients coded as "Do Not Announce".

Where possible, the following patients presenting with sensitive issues were also excluded by the organization where the patient received care (in advance of the data file being submitted to the vendor, Malatest):

- Patients who presented with confirmed or suspected sexual assault/abuse, elder abuse, or domestic violence;
- Patients who underwent a therapeutic abortion; and
- Patients deceased after discharge.

The sample plan was developed with completion targets at the facility level to achieve a margin of error (MOE) no greater than ±9% at the 95% confidence interval (CI). Individual health authorities were given an initial draft sample plan and Interior, Fraser, and PHSA opted to increased their sample for specific facilities to achieve lower target MOE (< ±7% or lower).

See also FAQs on Sampling and Survey Weighting

Samples of respondents in the left without being seen (LWBS) group were selected to be administered a reduced version of the survey questionnaire, with the objective of obtaining 40 surveys per health authority. At the end of the survey administration period, 281 survey completions had been obtained, with the final number of surveys obtained for each health authority ranging between 39 and 55 valid completions.

See technical report for more information on the Left Without Being Seen (LWBS) group.

The Data Collection Process

For the Emergency Department 2018 survey, data was collected via the following process:

1. **Data Submission:** Twice per month for the 3 month data collection period, Emergency Departments securely sent the selected survey vendor records of patients discharged from their facility. The sample data elements included with every patient record is included in the Data file Submission Manual. The survey vendor generated a random sample of patients from the "universe" of eligible patient records submitted. Eligibility required that the records included valid mailing addresses and phone numbers.

2. **Patient Notification:** Prior to being contacted, patients were notified by mail within 1 week of receiving discharge records that they had been selected to receive a survey. The cover letter also provided a unique access code and URL for those who preferred to complete the survey online.

3. **Survey Administration:** The surveys were then conducted by phone as an interview or self-completed online. All phone-based surveys were completed using computer-assisted telephone interview (CATI) and used standardized interview scripts and prompts (see survey next page).

Sampling Level	Data Collection Method
Facilities with response rates > 60% (up to 10% higher than the average)	• Eligible patients received two calls • Even after completion targets were met, the survey vendor will continue to accept completions via calls to their 1-800 number and online
Facilities with response rates between 40-59.9% (within ±10% of the average)	• Eligible patients received a minimum of three calls • Even after completion targets were met, Malatest will continue to accept completions via calls to their 1-800 number and online
Facilities with response rates <40% (more than 10% below the average)	• Eligible patients received a minimum of five calls • Even after completion targets have been met, Malatest will continue to accept survey completions via calls to their 1-800 number and online
Census Sites	• All eligible patients discharged from units with less than 350 unique discharges were provided the opportunity to participate in the survey; each patient will receive a minimum of five calls

All surveys, both online and phone, were available in the following languages:

English	Chinese	Punjabi	Korean

French	Spanish	German	Vietnamese

4

Data Collation: Patients’ survey responses were entered into a secure database and collated by the survey vendor. As noted, aggregated results and reports were provided to individual hospitals, health authorities and the province in May 2019.

Privacy Considerations

The information collected from patients on admission and given to the survey vendor for the purposes of conducting the survey included personal information required to conduct the survey (e.g., discharge date, phone number, and mailing address). Patient information that is used, disclosed, and retained for purposes of conducting Patient Experience of Care Surveys are statistical in nature; this means that results cannot be directly used to affect the treatment of a specific patient.

The survey vendor was required to demonstrate compliance with the BC Freedom of Information and Protection of Privacy Act and continues to be subject to the independent oversight of the BC Information and Privacy Commissioner. A Privacy Impact Assessment (PIA) for the project was completed and approved by the Health Information and Privacy Operations Committee of BC (HIPSOC) on September 10, 2017, and an onsite audit of the survey vendor's operations is conducted annually to review the way personal health information of BC patients is managed in each survey project. In addition, the survey vendor is contractually obligated to fulfill its obligations under BC's Privacy Protection Schedule.

In accordance with BCFOIPPA, which is a notification regime, throughout the time that the survey was being conducted, signs were posted in each of the 80 hospitals to advise patients that they may be selected to complete a survey. These posters fulfilled four purposes:

- Informing patients about the survey and the timeframe;
- Providing contact information, if patients have questions;
- Providing a mechanism for patients to "opt out"; and
- Providing information about the use and protection of the personal information of patients under BC's Freedom of Information and Protection of Privacy Act (BC FOIPPA).

In addition, a letter was mailed to each patient who was randomly selected from all hospital discharges that included specific information about the protection of personal information under BC FOIPPA, as well as contact information, if patients had questions about the survey, or wished to be removed from the survey contact list (see above for sample letter).

Response Rate

Response rate, along with the margin of error, is one of the most commonly used indicators to gauge the quality and accuracy of survey data.

The Emergency Department 2018 Survey was a mixed mode survey. Patients completed the questionnaire over the phone or online. In total, 64.3% of respondents completed the questionnaire by phone and 9.5% completed it online. Approximately 26.2% of respondents completed the questionnaire using both survey modes over multiple sessions.

The table below shows the response rate for each health authority. Response rates were calculated by dividing the number of completions over the valid sample (excluded ineligible patients).

The overall response rate for the Emergency Department 2018 Survey was 35.9%

Health Authority	Response Rate
Fraser	34.3%
Interior	38.8%
PHSA	37.9%
Providence	33.1%
Island	37.8%
Vancouver Coastal	34.0%
Northern	33.7%

See *Appendix G: Survey Disposition Results* in the Technical Report for response rates at the facility and unit levels

Survey Accuracy

An estimate of survey accuracy is the margin of error. The margin of error indicates the imprecision inherent in survey data. Margin of error is inversely related to the sample size used to draw inferences about the larger population. In general, larger sample sizes result in lower margin of errors and a smaller margin of error indicates the survey results were more precisely measured. A margin of error of ± 5% or ± 8% is considered good and acceptable, respectively.

The target survey completions and sampling methodology for the Emergency Department 2018 survey were designed to achieve a good to acceptable margin of error at the unit level. The Emergency Department survey had a ± 0.8% margin of error at the provincial level. The margin of error at the health organization level ranges from ± 1.8% to ± 5.2% (see table below).

Health Organization	Margin of Error
Fraser	± 1.9%
Interior	± 1.5%
PHSA	± 4.4%
Providence	± 5.2%
Island	± 2.0%
Vancouver Coastal	± 2.4%
Northern	± 1.8%

See *Appendix G: Survey Disposition Results* in the Technical Report for margin of error at the facility and unit levels

See also FAQs on Survey Accuracy and Response Rates

Response Categories

Response categories are the choices provided to respondents when asked a close-ended question. The choice of response categories and the number of scale points can affect how precise respondents rate their opinions and experiences. In general, more ordered response categories or a higher number of scale points allow finer distinctions to be made between patients' reported experiences and outcomes (i.e., higher degree of measurement precision). The associated increase in response variations also allows relationships between questions or dimensions to be examined to a greater extent. In doing so, results provide a better opportunity to detect changes and differences. However, if patients cannot reliably decide between two scale-points or the differences are not clinically meaningful, having additional response categories increases respondent burden and can add noise to the data, thereby increasing the amount of measurement errors.

The Emergency Department survey mostly adapted four-point scales without a neutral category (e.g., neither agree or disagree), with a mix of dichotomous questions and a 10-points rating scale for outcome related questions. Responses categories and number of scale points for the ED survey were determined using rigorous testing and validation processes, including cognitive interview and pilot studies that examined the scale reliability and validity of responses.

Valid and Non-Valid Responses

A response is considered "valid" when respondents select a response category that clearly states or reflects their opinion on a question (e.g., Never, Usually, Sometimes, Always). Valid response refer to the number of patients who provided a valid answer for the question and are used to calculate the valid percent. Responses such as "don't know" and "not applicable" and missing responses due to skip patterns are considered non-valid responses. "Don't know" is considered a non-valid response when calculating top-box scores as a "don't know" response cannot be classified as a positive or non-positive opinion or experience.

A Non-Valid response count refers to number of patients who did not provide a valid response (i.e., select from the valid response options) and answered "don't know", "not applicable", "prefer not to answer" to the question.

From the valid responses, a valid percentage is the percentage of responses (%) based on the total valid responses for a question or dimension. The valid percent column is arguably the best statistic for reporting purposes as it excludes those for whom the question was not applicable, and those who weren't sure of or didn't know the answer to the question.

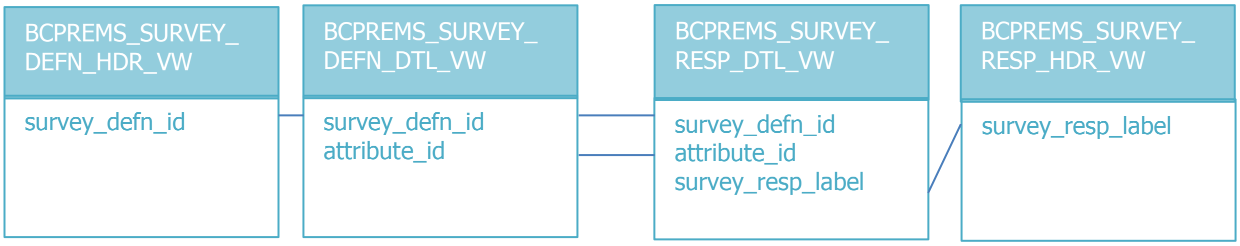
See *Appendix C 2018 Emergency Department Sector Survey Codebook* of the technical report to determine which response categories are valid responses and which are non-valid responses

Database Views in Healthideas

Healthideas normalized the vendor supplied individual data files into database objects designed for data security, storage efficiency, and scalability. Four database views were created from these objects for analysis purposes.

Database View	Description
BCPREMS_SURVEY_DEFN_HDR_VW	Contain information about the survey such as the survey sector and survey version.
BCPREMS_SURVEY_DEFN_DTL_VW	Contain information about the survey question such as the question unique identifier, question label, and the dimension the question belongs to.
BCPREMS_SURVEY_RESP_HDR_VW	Contain information about the respondents.
BCPREMS_SURVEY_RESP_DTL_VW	Contain response to a given question in the survey.

The four database views can be linked to each other with the following key columns.



Data Dictionary

There are two data dictionaries for this sector survey:

1) the data dictionary in Healthideas WIKI; and
2) the survey vendor supplied codebook.

The Healthideas data dictionary can be used as the primary data dictionary and describes the columns in the four database views developed for analysts. Users can then query the database view SURVEY_DEFN_DTL_VW to identify the labels for the survey questions and their response categories. The database views also contain information on which response categories are valid responses and which are non-valid responses.

See *Appendix C 2018 Emergency Department Sector Survey Codebook* of the technical report

The survey vendor supplied codebook provided the same information but there are some important differences between the two. Survey descriptors such as survey sections and variable names for each sector survey have been standardized into a common format when they were transferred to Healthideas. These meta data are standardized to facilitate the secure and efficient storage of multiple sector surveys. The codebook and questionnaire prepared by the survey vendors use the original variable names, instead of the standardized variable names. To find out the original variable names, a survey layout mapping document is available to map the Healthideas labels back to the original survey descriptors used by the survey vendor.

Missing Values for Dates

Missing values can be denoted implicitly as NULL values or explicitly with a special response value (code). In the Healthideas databases, all dates have been standardized into MMDDYYYY format. Dates that do not conform to this format, such as patients where no information on admission date exists or patients with incomplete dates, (e.g., only the month or year of admission is available) are shown as NULL values in the database.

"Partial" and "Complete" Surveys

While the exact definition of a complete survey varies depending on the sector and survey tool used, generally speaking, a partial survey means the respondent did not answer all questions. For example, if there were 100 questions, the patient only answered 75.

The Emergency Department reports include responses from partially complete surveys, answers from respondents who completed at least 50% of the survey. From the perspective that every patient's voice counts, each response, answers from patients who partially completed the survey should be included. From the perspective that every patient's voice counts, each response, including answers from patients who completed only one question should be included.

From an analytical standpoint, there are methodological challenges as to whether to include or exclude partially completed surveys. The challenge stems from the unknown (unobserved) systematic differences between patients who completed the entire questionnaire versus those who answered only some of the questions. The extent to which these differences cannot be adjusted or accounted for can bias the estimates. The number of non-missing responses within a survey instrument is also important when calculating summary scores of standardized instruments such as the VR-12. Standardized instruments often have strict guidelines on the minimum number or percentage of answered questions for a scale before summary scores can be calculated. For these reasons, in the report, missing VR-12 responses were imputed using multiple imputation (see Key Driver Analysis section).

While there is no right or wrong approach on how to handle partial completes, here are some guidelines:

- To replicate the numbers in the published reports, include partially completed surveys
- To examine potential differences between wholly complete and partially complete surveys, include partially completed surveys by treating them as a separate sub-group in the analysis
- To replicate the summary scores for a standardized instrument, consult the instrument developer's scoring manual and follow the recommended scoring algorithm and procedure

Scoring

Top Box Scoring

A Top-box score is the percentage of respondents who selected the most positive response category to a survey question. To facilitate interpretation of survey results and comparison across questions, survey responses are often standardized as a percentage of the most positive answers.

For individual questions, "Top-box" answers are defined as the most positive response category to a survey question regardless of the response categories. Results are easier to compare when they are all scored in this way, since there is less variation in interpretation of what constitutes a "good score."

Top-box score is calculated by dividing the sum of the most positive response over the sum of all valid response. The result is multiplied by 100 to transform it into a percentage.

$$\text{Top} - \text{box Score} = \frac{\sum \text{most positive response}}{\sum \text{valid response}} \times \ 100$$

For dimension and sub-dimension scores, the percentage of top-box responses for each question is first calculated separately and then averaged for dimensions/sub-dimensions that make up of multiple questions. In other words, dimension scores are calculated using an "average of the average" approach.

$$\text{Dimension Score} = \text{Average}(\text{top} - \text{box score for all questions})$$

An alternate method can be used, as in other sector surveys (e.g., Mental Health and Substance Use survey 2010/11), where the top-box dimension or sub-dimension score is calculated by treating all top-box responses and all valid responses for all the questions as one combined question (i.e., the "grand average" approach). The top-box scores are then calculated in two steps. First, the two total scores are calculated for each survey respondent. The first total score (top-box totals) consists of the sum of all "top-box" values for questions corresponding to each dimension. The second total score (valid response totals) consists of the sum of all valid responses for questions corresponding to the same dimension. Depending on the particular grouping or aggregation that was required, the top-box totals are divided by the valid response totals to obtain a top-box dimension score. The result is multiplied by 100 to transform it into a percentage.

$$\text{Dimension Score} = \frac{\sum \text{topbox response for all question}}{\sum \text{valid response for all question}} \times \ 100$$

Computationally, the "grand average” approach is more intensive when data is weighted. Conceptually, the two methods are different ways of calculating the same thing. In practice, unless the number of valid responses for each question varies significantly, the final dimension scores estimates are close enough that it does not make any practical differences in which approach to use.

Key Driver Analysis

A key driver analysis was conducted to identify which patient experiences are most likely to affect patients' overall experiences of care. Specifically, the key driver analysis utilized a Structural Equation Modeling (SEM) approach to identify dimensions (key drivers) and corresponding PREM items that reflect aspects of care strongly associated with four global rating indicators of experiences of care in the Emergency Department.

Global Rating Questions

EDPEC29. *Using any number from 0 to 10, where 0 is the worst care possible and 10 is the best care possible, what number would you use to rate your care during this emergency department visit?*

EDPEC30. *Would you recommend this emergency department to your friends and family?*

BCED9. *Overall, on a scale of 0 to 10, do you feel you were helped by your visit to the emergency department? Please answer on a scale where 0 is "not helped at all" and 10 is "helped completely."*

BCED10. *On a scale of 0 to 10, what was your overall experience with your emergency department? Please answer on a scale where 0 is "I had a very poor experience" and 10 is "I had a very good experience."*

The key driver analysis process involved the following 4 steps:

1) Treatment of missing data

- Missing data occurs when an individual's valid response to a survey question is unknown. Reasons for missing data include an individual not knowing what response to give, an individual preferring not to answer, or a question not being administered to an individual or group (e.g., because the question is not applicable to all people). In a multivariable analysis, missing data are treated by either excluding individuals who do not have a valid value for any of the variables included in the analysis, or by implementing statistical methods to accommodate the missing values. The exclusion of individuals leads to biased estimates in inferential analyses (e.g., the results may no longer be representative of the same population). It is therefore recommended to use statistical methods to adjust for potential biases associated with missing data. Multiple imputation is a statistical method that is widely recommended as a robust approach for minimizing biases due to missing value in multivariable analysis. Prior to running analyses on the 2018 Emergency Department data, *multiple imputation* was used as a method for estimating the values of missing data for the respondents who completed more than 50% of the survey.

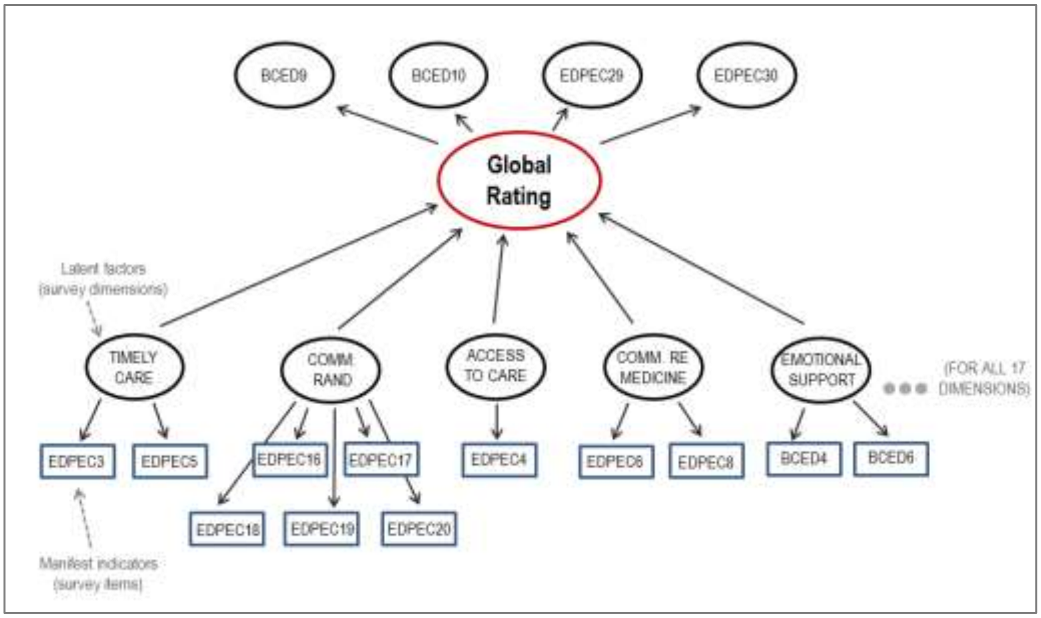
2) Factor analysis

- Prior to running the analyses, 16 survey dimensions were identified based on previous validation work with the EDPEC items and categorization of new items for the BC ED Survey. A total of 55 items were initially classified into these 16 dimensions. Factor analyses were conducted using the MPLUS software (version 8) to confirm the plausibility of 16 dimensions as correlated latent factors. A probit link function and mean and variance adjusted weighted least squares estimation was used to fit the model while accommodating the ordered categorical item distributions. The 16-factor model resulted in a reasonable fit. However, the item CONT12 did not correspond well with the managerial continuity domain. Consequently, this item was separated out as a separate dimension, named "CONTACT". For a complete list of the dimensions and items within each dimension, consult the Technical Report.

3) Structural equation modeling

- Structural Equation Modeling (SEM) is a statistical technique that combines psychometric factor analysis and multiple regression analysis in a single statistical model. SEM was chosen as an appropriate methodological approach for identifying key drivers because of (1) its ability to account for correlations between survey items and (2) its ability to control for the effects of multiple dimensions on global ratings in order to tease out the unique contributions of each dimension.

  The SEM analysis was implemented by specifying the 17 latent factors representing the survey dimensions as predictors of the four global ratings indicators, which were represented as an additional latent factor. This structure is illustrated in the following path diagram:



4) Variable importance analysis

- The Pratt index[1] was used to identify key drivers from the SEM analysis. This index can be interpreted as quantifying the relative contribution of each dimension to the total explained variance of the global ratings (the proportion of explained variance in the global ratings attributable to a given dimension). In other words, it helps to answer the question: "What explains why one patient rates their ED visit highly while another rates their ED visit poorly?"

---

[1] Thomas, D. R., Hughes, E., & Zumbo, B. D. (1998). On variable importance in linear regression. *Social Indicators Research, 45*, 253-275.

Qualitative Comments

In addition to the close-ended questions, Patients were asked to provide narrative comments at the end of the questionnaire in response to the question, **"What is the most important change we could make to improve patient experiences in BC Emergency Departments and ambulance services? We welcome your additional comments."**

For the purposes of the 2018 Emergency Department survey, open-text comments were transcribed verbatim if the survey was completed over the phone and are written exactly as entered if the survey was completed online. All comments appear verbatim in the data set, with no corrections for grammar or content, although an personal identifiers are masked (XXXX). The survey vendor then adapted the coding scheme used in the *2016/17 Acute Inpatient Sector Survey*. The coding scheme was then refined by, and approved by, the BCPCMWG. The coding scheme, which is presented in Appendix A of the Emergency Department 2018 Technical Repot, has 44 individual themes (the *2016/17 Acute Inpatient Sector Survey's* coding scheme had 38 themes). For each theme, valence codes were assigned depending on whether the theme-specific comment was positive, negative, neutral, or positive and negative. Prior to being included in the facility level reports and the data sets for Healthideas, the survey vendor reviewed all comments to remove identifiers that could reveal the identity of the patient, doctors, nurses, or other staff. Also, comments that were insensitive to specific racial or ethnic groups were adjusted so that the group was no longer identifiable. Narrative comments are included at the record level in Healthideas.

See Technical Report for additional details on coding themes and categories.

Open-text comments serve as a rich source of qualitative data to compliment the quantitative results of the survey. Open-text comments can be used to illustrate the human face of the data, to provide additional insight into what the survey results are demonstrating, and to point to areas not addressed in the survey that may be important to patients.

Peer Groups

Peer groups are useful for the assessment of survey results, particularly when it comes to performance improvement, as they allow comparison between similar facilities/reporting levels. Properly constructed, they convey the range of scores for similar units and facilities and serve as benchmark scores for comparing units or facilities results with their provincial and national peers. In the Emergency Department sector, the following peer groups were used:

Size of Facility	Other Peer Group
Large Medium Small Extra-Small	Youth NACRS Facilities

Size of Facility

The facility is classified into one of four type based on their annual patients volume,.

Peer Group	Definition
Large	Facilities with more than 40,000 annual patient visits
Medium	Facilities with between 20,000 and 39,999 annual patient visits
Small	Facilities with between 5,000 and 19,999 annual patient visits
Extra-Small	Facilities with fewer than 5,000 annual patient visits

Other Peer Group

A patient can be classified into this additional peer group:

Peer Group	Definition
Youth	Any patient who is less than 18 years of age.
NACRS Facilities	One of 29 National Ambulatory Care Reporting System (NACRS) emergency departments in BC

FAQs: Sampling and Survey Weighting

How were patients chosen to participate in the survey?

See Technical Report for additional information on how the sample was selected and prepared (e.g., de-duplication).

The 2018 Emergency Department Sector survey used a disproportionate stratified random sampling design to select patients. Patients were randomly selected from the 108 participated emergency department at the facility level to complete the survey voluntarily. Patients were only contacted and invited to participate once during the surveying period (deduplication was carried out to ensure that patients would not be invited to participate more than once during the survey period). To ensure statistical accuracy of the results, all patients from facilities with fewer than 350 discharges between January 1 – March 31, 2018 were invited to participate (i.e., a census approach). In all other facilities, patients were randomly sampled with the aim of achieving a balanced number of completions across the 3 months of data collection. All other units were sampled twice a month (patients discharges from 1st to 15th and 16th to the end of each month) for total of 6 times over a 3 month period. Completion targets were set to ensure the facility margin of errors to be within ± 9%. At the request of some health authorities, the number of patients to sample for some facilities in Interior, Fraser, and Provincial Health Services Authority were set to achieve a margin of errors as low as ± 4.5%.

In stratified random sampling, the patient population is divided into two or more groups (strata) according to one or more common characteristics before randomly selecting patients from each stratum. For this survey, the strata is the facility. In disproportionate stratified random sampling, the numbers of patients recruited from each stratum is not proportionate to the total size of the patients population. The sampling plan was designed to obtain more precise information on the smaller subgroups by over sampling smaller units and under sampling larger units (i.e., disproportionate stratified random sampling). Compared to simple random sampling, stratifying tends to reduce sampling error and ensures a greater representation from subgroups. When members within the strata are more similar (homogeneity) than members between strata (heterogeneity), survey estimates can be as precise (or even more precise) as simple random sampling.

What are survey weights?

Survey weights are used to make the analysis sample representative of the target population on key characteristics. Key characteristics may include organization level attributes such as discharge volume and facility type or demographic characteristics such as age group and gender. Survey weights, or the inverse probabilities of selection for each observation, allow us to reconfigure the sample as if it was a simple random draw of patients that is representative of the total patient population to yield accurate estimates.

Why does the BCPCM Working Group use survey weights?

Survey weights are used to make the analysis sample representative of the target population on key characteristics. Surveys are often designed to obtain more precise information on smaller subgroups by over sampling smaller units and under sampling larger units. As a result, the way key characteristics, such as discharge volume, are distributed may differ at the sample level from the way they are distributed in the population. For example, a sample at the facility-level may consist of 50% patients from unit A and 50% from unit B, when, in actuality, unit B makes up of 75% of the facility's discharge volume in the population. This disproportionate sampling introduces bias into the population estimate you may obtain from your sample because statistical procedures will give greater weight to people you over-sampled and less weight to those you under-sampled.

The BC Patient-Centred Measurement Working Group corrects for these biases with post-stratification survey weights. Survey weights, or the inverse probabilities of selection for each observation, allow us to reconfigure the sample as if it was a simple random draw of patients that is representative of the total patient population. Without weighting the data, patients' responses from over or under-sampled units will be given more or less weight in their answers than they should, resulting in biased population estimates.

The survey weights available for the 2018 Emergency Department survey are post-stratification weights that reweight the sample responses to match the population distribution in terms of **discharge volume.**

When does the BCPCM WG apply survey weights?

Depending on the level of analyses and research questions, the BC Patient-Centred Measurement Working Group applies weights so responses are to be representative of the patient population in terms of discharge volume at the unit, facility, and health authority levels.

When working with the PCM sector surveys, if the analytic questions involve comparing results across organizational units (e.g., unit, facility, & peer group), survey weights are recommended to ensure the analyses yield estimates that are less likely to be biased. If the analysis focuses on findings from only a single organizational unit (i.e., the unit level), consider applying additional individual level weights to account for demographic differences due to sampling as the supplied survey weights for PCM data often only account for differences at the organizational level. If the goal is to estimate causal effects and examine relationships between variables, then there are situations that call for the use of weights and situations that don't. Consult with a statistician for recommendations.

What are post-stratification weights?

Post-stratification weights are survey weights that are computed after you have collected all your data. The stratification part occurs when the patient population is first divided into two or more groups (strata) according to one or more common characteristics before randomly selecting patients from each stratum. For the 2018 Emergency Department survey, strata were determined based on patient discharge volume at the facility levels.

Do survey weights handle non-response?

Non-response occurs when a patient invited to participate in a survey does not answer one or all survey questions irrespective. Non-response bias is the difference in the results for those who responded versus those who did not respond (e.g., they are unwilling or they are unable to) for a survey.

Post-stratification weights can indirectly adjust imbalances with respect to discharge volume in the sample due to non-response. Survey data can be re-weighted to bring the sample discharge volume more closely into line with the population discharge volume. This approach is known as non-response weighting and post-stratification weights are one such method that takes into account non-response indirectly. Another approach to survey non-response is data imputation and a model-based approach.

What survey weights are available?

The 2018 Emergency Department survey dataset contains two post-stratification survey weights – expansion weights and normalized weights. Expansion weights are weights that sum up to the number of patients discharges at the population. Normalized weights are rescaled expansion weights wherein the weights sum to the sample size. Both set of weights reweight the sample to match the population distribution of discharge at the facility, health authority, and provincial levels.

Survey Weight	Description
WEIGHTS	The weights were calculated in two steps. First, an initial weight that reweighted the sample to match the population discharge volumes at the facility level. This weight included discharge volumes from facilities that did not receive their facility level report due to various reasons, such as no patients completing a survey from that facility (i.e., miscellaneous facilities). In the second step, the initial weights were adjusted so that the sample match the population discharge volume at the health authority level to account for any remaining differences between the sum of all weighted facilities and that health authority total discharge volume. This adjustment only affects health authorities where there was no survey completion from the miscellaneous facility but in the population, we have identified at least one miscellaneous facility, and therefore the initial weight cannot account for these discharges because we do not have any completion for them. Since all facilities were accounted for within health authorities after the second adjustment, no further adjustments/weight calculations were necessary. The final survey weight was used in the production of the annual reports and storyboards.
WEIGHTS_NORMAL	The normalized version of WEIGHTS so that the weights sum up to the sample. When running statistical procedures and comparisons, the normalized weight should be used when the software does not consider the survey sampling design information.

How were the survey weights calculated?

The survey post-stratification weights were calculated in two stages using R *survey* package. First, to account for the population discharge volume of the participated facilities, and second, to account for facilities not in the final sample due to noncompletions or other reasons so that the unit totals sum up to the total number of patients in each health authority. For the purposes of weighting, discharges include all patients in the population frame regardless of their eligible status. To calculate the weights needed to adjust the sampling distribution to match the population discharge volumes, control tables with the discharge volume by facilities within each health authority were used. Virtually all facilities were accounted for within health authorities (with the exception for two facilities within Interior Health, one of which had no discharges and the other had only two discharges but no contact information, so no survey administration was possible). A very slight adjustment to the weights was made to determine the health authority level weights. The two discharges not accounted for in the facility level weights for the one health authority were not enough to make any appreciable difference between the weighted results using facility-level or health-authority-level weights, therefore the health authority level weights were used for all analyses. As provincial discharge volume is composed of the sum of all health authorities discharge volumes, weighting at the health authority and levels were sufficient to account for the distribution of discharges at the provincial level.

A review of the distributions of the survey completions revealed that the survey sample for the LWBS group did not match the distribution of the sample universe by age, facility, peer group (facility size), or health authority. The imbalance by health authority, peer group, and facility was not surprising given the restriction of the survey samples to about 40 survey completions per health authority. Only 39 surveys were obtained for the FHA, which represents over 45% of all LWBS cases in the survey frame, with fully 49 surveys for the PHSA (BC Children's Hospital), which represents only 14% of all LWBS cases. It may also be noted, that at client request, over-sampling of elderly LWBS cases was undertaken at one of the health authorities.

The decision was made to weight by individual age group (10 groups) rather than just the three aggregated age ranges used for analysis (0-17, 18-64, 65+). The design effect[2] associated with this data weighting scheme is modestly higher compared with using the three aggregated age ranges, however this scheme does ensure that within the broad 18-64 age group, the 18-30 and 30-45 age ranges (which were notably under-represented in the unweighted sample) are appropriately represented in the weighted sample; and the 0-6, 65-75, and 75-85 age ranges are not disproportionately over-represented within their broader age groups. Thus, when the three broad age groups (0-17, 18-64, 65+) are reported on individually, they should theoretically provide a more accurate representation of the results than without the weighting (notwithstanding the decrease in the effective sample size due to data weighting).

What is an expansion weight?

---

[2] The sampling design effect is a measure of the extent to which sampling error is increased due to the application of data weights to correct for biases in the survey sample and the resulting decrease in the effective sample size). computed as *deff = sample size n * (sum of the squared weights) / (square of the summed weights)*. Sampling efficiency *= 1/deff*. Effective sample size = *sample size n * sampling efficiency*.

Expansion weights are survey weights where the sum of weights adds up to the population count. A weight is a value assigned to each case in the data. The value indicates how much each case will count in a statistical procedure. For example, a weight of 10 means that the case counts in the dataset as 10 identical cases whereas a weight of 1 means that the case only counts as one case in the dataset. Survey weights can be and often are fractions, but are always positive and non-zero. Please note that software that does not consider survey sampling design often equates the sum of weights with the number of observations.

In general, using the expansion weights available in such software results in the underestimation of variance and in too many results being declared as significant when conducting statistical testing.

What is a normalized weight?

Normalized weights are rescaled weights where the sum of weights equals the sample size. It considers the survey weights, but not other aspects of the sampling design such as stratification, clustering, or calibration. When the weighted number of patients is very different from the unweighted number of cases, software that does not account for the survey sampling design will not be valid because it over-estimates the number of cases used in the tests as the software associates the sum of the weights with the number of observations (or the effective sample size). This generally results in an underestimation of variance and in too many results being declared as significant. Normalized weights partially address the problem with expansion weights by keeping the sum of weights the same as the sample size.

Are weighted results rounded?

Survey weights are often fractional numbers and some degree of rounding is involved in the estimates. The statistical programming language R was used to compute the survey weights. R uses un-biased rounding (rounding half to even) where real numbers are rounded to the nearest integer, except where the decimal places are exactly 5. In these cases, the statistical programming language R rounds to the even integer by default (e.g., 82.65% is rounded to 82.6%). For weighted frequency count, it is customary to round them to the nearest whole number when reporting. As a result, small discrepancy of .1 percentage point is to be expected when comparing the percentage of combined individual response categories against the sum percentage of multiple categories.

FAQs: Response Rate and Survey Accuracy

How precise are the survey estimates?

Numbers are rounded to avoid reporting insignificant figures. For example, it would create false precision to express a top-box score as 90.60000 % (which has seven significant figures) because the questions were never designed or validated to measure patients reported experience and opinion to that degree of precision.

For the purpose of reporting, most survey results, including top-box scores or subgroup averages are considered to have a level of precision of up to one decimal place and are stored internally up to the precision allowed by the software. Given the measurement precision of the survey question, reporting more than two decimals of precision is not recommended as the original survey questions are unlikely to measure patients' experience and opinions accurately to two or more decimal points.

What is non-response bias?

Non-response bias is the bias that results when non-respondents differ systematically in meaningful ways from respondents. The result is that the survey sample often doesn't reflect the population they are meant to represent very well.

Most surveys suffer from non-response bias that may affect the quality of a survey and how accurate its estimates are. When patients who were selected in the random sample are unwilling or unable to participate in the survey, they are said to provide a non-response to the questionnaire (i.e., are non-respondents). When patients decline to answer a particular question, they provided "a non-response", resulting in missing data at the question level.

What does response rate tell us about non-response bias?

It is important to note that response rate is not necessarily a good indicator for non-response bias. A higher response rate, while desirable, does not mean the survey has smaller bias. Conversely, a low response rate does not by itself imply that survey estimates are biased. Instead, knowing whether responses from respondents and non-respondents differ in some systematic way is the best indicator of non-response bias. One way to assess the impact of non-response is to compare known characteristics known to be related to survey responses between these two groups to see if they differ. Information on non-respondents might come from previous sector surveys or external administrative data sources (e.g., health records or para-data files). The degree of non-response bias may also differ depending on the reasons of non-response (e.g., refusal, non-contact, technical problem). Non-respondent bias is often difficult to assess because of a lack of information from external sources to compare their characteristics against those of the respondents.

What is method bias?

Systematic differences in survey responses obtained from respondents who completed the survey in different survey modes are part of a class of bias called method bias. Method bias is broadly defined as any bias that results from the choice of survey method. Method bias happens when variations in responses are associated with the method (or survey instruments) rather than the actual opinions or reported experience of the respondents that the survey attempts to measure. The bias can occur because of the way the items or questions are phrased, the way in which they're asked, or the audience to which they're asked (e.g., self-report versus proxy respondents). This may include differences related to survey mode (e.g., phone, mail, or online survey), response format (Likert-scale versus multiple responses), scale range (3-point vs. 5-point scale), positive or negative item wording, or the language in which the survey is conducted.

Method bias is one of the main sources of measurement error in mixed-mode surveys. The method introduces "noise" variations in responses that contaminate actual differences and variations in patient's opinions or experiences.

For standardized instruments, method bias and method variance can be accessed via psychometric techniques such as confirmatory factor analyses or item response theory.