A Goal-Driven Methodology for Developing Health Care Quality Metrics

Carlos Villar Corrales

Directed by:

Daniel Amyot and **Dominique Ferrand**

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of

Master of Science in Electronic Business Technologies

uOttawa

E-Business Technologies
Faculty of Graduate and Postdoctoral Studies
University of Ottawa

Abstract

The definition of metrics capable of reporting on quality issues is a difficult task in the health care sector. This thesis proposes a goal-driven methodology for the development, collection, and analysis of health care quality metrics that expose in a quantifiable way the progress of measurement goals stated by interested stakeholders. In other words, this methodology produces reports containing metrics that enable the understanding of information out of health care data. The resulting *Health Care Goal Question Metric* (HC-GQM) methodology is based on the Goal Question Metric (GQM) approach, a methodology originally created for the software development industry and adapted to the context and specificities of the health care sector. HC-GQM benefits from a double loop validation process where the methodology is first implemented, then analysed, and finally improved. The validation process takes place in the context of adverse event management and incident reporting initiatives at a Canadian teaching hospital, where the HC-GQM provides a set of meaningful metrics and reports on the occurrence of adverse events and incidents to the stakeholders involved. The results of a survey suggest that the users of HC-GQM have found it beneficial and would use it again.

Acknowledgments

I would like to thank the following people and organizations who have helped me to successfully complete this thesis:

- To NSERC and CIHR, for the funding provided through their Collaborative Health Research Project on *Performance Management at the Point of Care: Secure Data Delivery to Drive Clinical Decision Making Processes for Hospital Quality Control*.
- To the University of Ottawa, for the scholarship they have given me.
- To the personnel of the Ottawa Hospital Research Institute and of The Ottawa Hospital for opening their doors to my research, for collaborating with me, and for providing me with cases studies.
- To my family, for always supporting me and believing in me.
- To Charlie, for her patience, comfort, and help editing this document.

Finally, I would like to give special thanks to my supervisors, Daniel Amyot and Dominique Ferrand, for all the time spent on my thesis, providing me with great guidance.

Thank you.

Publications

Ferrand, D., Amyot, D., Villar Corrales, C. (2010). Towards a Business Intelligence Framework for Healthcare Safety, Journal of Internet Banking and Commerce, vol. 15, no. 3, available at http://www.arraydev.com/commerce/jibc/2010-12/Ferrand.pdf.

Ferrand, D., Amyot, D., Villar Corrales, C. (2010). Towards a Business Intelligence Framework for Healthcare Safety, Proceedings of the International Conference in eCommerce and ePayment ICeP'10, Gdansk, Poland, September 25-27, pp 71-80.

Table of Contents

List of Figures

List of Tables

List of Acronyms

Acronym	Definition
AE	Adverse Event
AEMS	Adverse Event Management System
AMS	Antibiotic Management System
BI	Business Intelligence
CMM	Capability Maturity Model
DW	Data Warehouse
GQM	Goal Question Metric
HC-GQM	Health Care Goal Question Metric
HCV	Hierarchical Cumulative Voting
ICU	Intensive Care Unit
IQIP	International Quality Improvement Program
OHRI	Ottawa Hospital Research Institute
PATH	Performance Assessment Tool for quality improvement in Hospitals
PSLS	Patient Safety Learning System
PSM	Practical Software and Systems Measurement
QCIPA	Quality of Care Information Protection Act
TOH	The Ottawa Hospital

Chapter 1. Introduction

In Canada, of the approximately 2.5 million patients that are admitted every year to hospitals, about 185,000 experience undesirable negative outcomes caused by medical care. About 70,000 of those are deemed to be potentially preventable [Baker et al., 2004 p1]. One of the top priorities of health care is the provision of "high-quality" medical attention to the people who need it [Behnam et al., 2009 p1]. This objective however, is not completely achieved, since thousands of patients experience the consequences of *adverse events* (AEs). Minimizing the number of these AEs could incrementally contribute to the quality of care, which may translate into improving patients' safety. Medical centers in Canada and in many other countries are currently striving to fulfill this goal [Forster et al., 2004, p1]. The Canadian government realized the importance of such a goal and in the year 2002 budgeted $50 million over 5 years for the creation of the Canadian Patient Safety Institute [Baker et al., 2004 p1]. Moreover the reporting of AEs "is becoming a legal obligation in many provinces and states" [Behnam et al., 2009 p1]. Ontario and Quebec are a close example of such regulation [Canadian Medical Protective Association, 2009 p6].

The ability to quantify the quality of care, such as adverse events and other types of incidents, is an important element to improving patient safety. This quantification is predicated by measurements, which "…describe phenomena in terms that can be analyzed statistically" [Hulley et al., 2001, p37]. As a consequence there must be mechanisms in place to help measuring quality in the health care sector.

This thesis focuses on providing a methodology that contributes towards the tasks of discovering metrics capable of quantifying elements of quality of care and of producing reports exploiting those metrics to enable the understanding of information out of health care data. This new methodology is obtained by drawing upon previous methodologies and experiences gathered in a study at the Ottawa Hospital in 2009-2010.

1.1. Concepts

This section provides readers with a definition of concepts needed to understand the research context. These concepts are referenced in forthcoming chapters.

Goal: As defined by World Health Organization, a goal is a "general objective towards which to strive." [WHO EMRO, 2004]

Patient Safety: Defined by the U.S. Institute of Medicine (IOM) as "the prevention of harm to patients" [Aspden et al., 2004]. The Patient Safety Network Web site of the Agency for Health Research and Quality (AHRQ) defines it as: "freedom from accidental or preventable injuries produced by medical care." [AHRQ, 2010]

Quality of Care: Defined by IOM as "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge." [Lohr, 1990]. Quality of care includes patient safety.

Adverse Events (AEs): Defined as "unintended injuries or complications that are caused by health care management, rather than by the patient's underlying disease, and that lead to death, disability at the time of discharge or prolonged hospital stays." [Baker et al., 2004 p1] They can be classified as preventable or non preventable, the former being the type of AE that brings most concern to health care organizations. AEs are also known as "medical errors". Further they can be defined as:

- "Any response to medical care in the hospital that is unintended, undesirable, and harmful to the patient" [McLamb and Huntley, 1967 p469]
- "An untoward or undesirable occurrence in the health care process, which has or potentially has some negative impact on a patient or patients and results or may result from some part of the health care process" [Walshe, 1998 p94]

Incident: "A specific unplanned event or sequence of events that has an unwanted and unintended consequence on the safety or health of people, property or the environment, or on legal or

regulatory compliance" [Safety Moment, 2010]. It should be pointed out that this definition contains the concept of AE. Therefore an AE is a specific type of incident.

Business Intelligence: "We define business intelligence (BI) as systems that combine data gathering, data storage, and knowledge management with analysis to evaluate complex corporate and competitive information for presentation to planners and decision makers, with the objective of improving the timeliness and the quality of the input to the decision process." [Negash et al., 2008 p176]

Dimensional Modeling: Technique used to obtain a dimensional model. "This type of model is very popular in data warehousing because it can provide much better query performance, especially on very large queries, than an entity-relationship model. However, it also has the major benefit of being easier to understand." [Ballard et al., 2006 p52]

Data: "Data are often viewed as the lowest level of abstraction from which information and then knowledge are derived. Raw data refers to a collection of numbers, characters, images or other outputs from devices that collect information to convert physical quantities into symbols that are unprocessed." [Wikipedia, 2010]

Data Warehouse: As defined by Oracle "A data warehouse is a relational database that is designed for query and analysis rather than for transaction processing. It usually contains historical data derived from transaction data, but it can include data from other sources. It separates analysis workload from transaction workload and enables an organization to consolidate data from several sources." [Lane et al., 2002 p42]

Measure: Defined as "the number or symbol assigned to an entity […] in order to characterize an attribute" [Fenton & Pfleeger, 1997]. For example, "10 adverse events per month" is a specific measure.

Metric: Defined by the Glossary of Software Engineering Terminology as "A quantitative measure of the degree to which a system, component, or process possesses a given attribute" [IEEE, 1990]. For example, the "total count of adverse events per month" is a specific metrics.

Indicator: "A variable with characteristics of quality, quantity, and time used to measure, directly or indirectly, changes in a situation and to appreciate the progress made in addressing it. It also provides a basis for developing adequate plans for improvement" [WHO EMRO, 2004]. For example, the "total count of adverse events per month, where the target is less than 5" is a specific indicator.

1.2. Research Context

One of the main goals of modern health care is to improve quality of care [Behnam et al., 2009 p1]. According to the U.S. Institute of Medicine [IOM, 2001 p3], this goal is considered a multidimensional construct and is encompassed by concepts of patient safety, equity, timeliness of care, efficiency, and effectiveness, among others. One of these concepts, patient safety, has been catalogued since 1999 as "one urgent care problem" [IOM, 2001 p2].

Today, in 2010, patient safety remains an unresolved issue that hospitals strive to find ways to improve. One way of reaching this goal is by reducing the number of preventable adverse events. These errors claim around 28,000 lives annually in Canada [Behnam et al., 2009 p3] and as many as approximately 98,000 in the USA, exceeding "feared threats as motor-vehicle wrecks, breast cancer, and AIDS" [Kohn et al., 1999 p1]. AEs are generally caused by "faulty systems, processes, and conditions that lead people to make mistakes or fail to prevent them" [Kohn et al., 1999 p2]. Without proper mechanisms in place to assess and more importantly report on these faulty systems, processes, and conditions, it is not likely that we can reduce the number of AEs, improve patient safety and, therefore, improve the quality of care.

Many methodologies have been developed in the past to identify and manage AEs: mandatory and voluntary reporting systems, chart reviews, malpractice claims analysis, prospective clinical surveillance, and patient observation are a few examples [Thomas and Petersen, 2003]. Although these mechanisms are able to identify AEs, without proper quantification of the collected information, it is difficult to assess any stage of progress, learn how processes perform, or obtain guidelines on how to reduce the number of AEs. By discovering *metrics* capable of quan-

tifying such phenomena, we are able to extract knowledge. This provides stakeholders with a powerful tool to perform meaningful discoveries that would not have been possible otherwise.

Metrics that provide quantification of collected information and therefore generating "intelligence" are not only necessary for the reduction of AEs but are also a mandatory tool for assessing quality of care in general. For this reason, and because AEs are one major kind of health care quality issues, the development of a methodology that allows the discovery of metrics in the context of health care quality is of the utmost importance.

1.3. Research Problem

The definition of metrics capable of reporting on quality issues is a difficult task in the health care sector. "Quality-of-care literature is full of discussions about performance measurement" [Donaldson, 1999 p7] and many quality models have been implemented or adapted for such purpose. Continuous Quality Improvement (CQI) and Total Quality Management (TQM) are typical examples of adapted methodologies [Arah et al., 2003]. These models are however quite generic, and the discovery of metrics is one process among the many processes to be performed. Furthermore, such models present "measurability" issues that should be addressed before applying them to the health care domain [Kahan and Goodstadt, 1999].

Many methodologies have been developed for the discovery and collection of AEs. These methodologies provide a pool of raw, factual data whose value is generally unknown. As mentioned by Michalowski, "hospitals are data rich, but information poor" [Trum, 2010]. What stakeholders want is to extract knowledge out of this data. Such knowledge comes in the form of reports containing aggregated and quantified information that can help them make decisions. However, the process by which these collected values are quantified and transformed in actionable information is missing. What this situation ensues is a gap between the data gathered and the need for or knowledge of its applicability. Therefore discovering and generating metrics that quantify data in such a way that it can provide stakeholders with "knowledge" is essential.

Many methodologies have emerged to address this existing gap. Examples of those are the International Quality Improvement Program (IQIP) and the Performance Assessment Tool for quality improvement in Hospitals (PATH) [IQIP, 2010; Veillard et al., 2005]. There are, however, many issues that limit the ease of implementation or applicability of such methodolo-

gies in the health care context. Some of them have been outlined in the literature and are presented below:

- It is common practice to take already developed metrics and use them in different contexts without a proper monitoring that ensures their performance. This brings as a consequence that metrics do not accomplish the expected levels of validity and reliability [Walsh, 2000 p51].
- There are concerns about confidentiality and non-disclosure of information [Thomas & Petersen, 2003 p63; IOM, 1999 p19]. This problem can translate into the unavailability of using information that resides in different sources, and therefore obstructing measurement purposes. These concerns are also linked to other issues such as care providers being afraid that information that is not "confidential" can be used against them.
- Metrics are generated following recommendations of "best practices" or through the "National Library of Health Care Indicators" assembled by the Joint Commission on Accreditation of Health Care Organizations (JCAHO). This practice uses "benchmarking" without taking into consideration the specific context problems that health care centers may have. Therefore, metrics might not be geared towards accomplishing specific goals and the measurement exercise loses focus.
- Many methodologies do not address the technology used for data collection and storing. This brings as a consequence that data is stored in spreadsheets (such as Microsoft Excel), poorly designed databases or even word processing documents (e.g., Microsoft Word). The possibilities offered by dimensional models and data warehouses, such as natural aggregations or dimensionality of the stored information, are not taken into account.
- Some metrics are not suitable for different populations [Donaldson, 1999 p2; Thomas & Petersen, 2003; Spertus et al., 2005]. Some metrics are valid for some patient populations while the same set is not valid for others. If these constrains are not properly introduced in the measurement approach, it becomes easy to forget about them and use the collected data in erroneous fashions.
- The fear of care providers to ruin their reputations or be engaged in lawsuits is another concern [Thomas & Petersen, 2003 p63]. As a result, care providers feel threatened and

may not participate in the measurement initiatives. This is counterproductive because care providers' input is necessary to develop precise measurement goals.

- "For the clinicians involved, focusing on adverse events to the exclusion of other things could be dispiriting and de-motivating." [Walsh, 2000 p51]
- The development of meaningful and reliable definitions of AE is difficult if they have to be used towards measurement because they might be discovered through several distinct methods that are not always quite reliable or valid themselves [Walsh, 2000 p51].
- Lack of clarity in who is responsible for the performance exercise [Donaldson, 1999 p21].

After considering the current situation, it can be stated that the research problem faced by this thesis is to address the existing gap between the data collected and the metrics needed to report on this data while targeting the reported shortcomings of existing methodologies.

1.4. Motivation

The fact that 70,000 patients suffer from preventable AEs in Canada is alarming. Moreover, AEs are just one of the many types of incidents that can occur in health care. In this context, the quantification of quality of care is essential for tangible improvements. Although "Medicine is experiencing an unprecedented increased focus on quantifying and improving the quality of health care" [Spertus et al., 2005 p1704], none of the reviewed methodologies addresses or takes into consideration most of the factors that limit the potential for developing metrics that can report on AEs, incidents, patient safety, and ultimately on the quality of care.

The main motivation behind this study is to help stakeholders address most of the issues related to the development of metrics described in Section 1.3. It is necessary to provide guidance to the process of metrics discovery and to obtain well-validated metrics capable of reporting on health care quality in general, and on AEs and other incidents and their consequences in particular.

Furthermore, what drives the core of this investigation is the learning process that will make it possible to develop a methodology guided by stakeholders' goals and objectives, so they obtain metrics that are meaningful to them and aligned with their goals. This will be accom-

plished by conceptualizing and assessing the experience gathered during this research. This methodology should be general enough to go beyond the context where it was created (measuring AEs and incidents at a teaching hospital) and be applicable to many processes and quality aspects of the Canadian health care system.

1.5. Research Objectives

The main objectives of this research are:

- *The formulation of a methodology that allows the development, collection, and analysis of health care quality metrics that expose in a quantifiable way the progress of measurement goals stated by the interested stakeholders*. This methodology is expected to combine aspects of other methodologies while avoiding some of the difficulties mentioned in Section 1.3. Consequently, it should provide health care professionals with guidance to finding a small, useful set of metrics that they can act upon. This is based on the hypothesis that a particular metric is only useful if it helps with the understanding of an underlying process or one of its resultant products. The methodology should indicate what information should be presented in the reports, why certain metrics are needed more than others, as well as for whom the reports and their metrics are intended. As a result, the methodology should present users with structured and organized reports containing the information required to assess quality related to the health care domain. The context in which the methodology is developed and evaluated is that of the measurement of processes, systems, and conditions that lead to the occurrence of AEs and other incidents. Once the methodology has undergone a validation process, it is expected that will have the potential to be general enough to be used towards other health care processes or areas.
- As a tool to support the methodology and the storing of the collected data, *the development of a dimensional model is also targeted*. A data warehouse (DW) and diverse data marts should be created considering the needs of the methodology. The reason to build a DW is that these structures are ideal for the presentation of information. In other words, they "support more fact-based decision making" [Kimball and Ross, 2002] and make it easier to access the right data at the right moment.

1.6. Research Method

The research method described in this section helped accomplish the research objectives. It uses an incremental and iterative approach. This means that each step was not addressed completely by working on it once, but by addressing it several times until a proper level of validation was achieved.

A visual representation of the research method is shown in Figure 1.

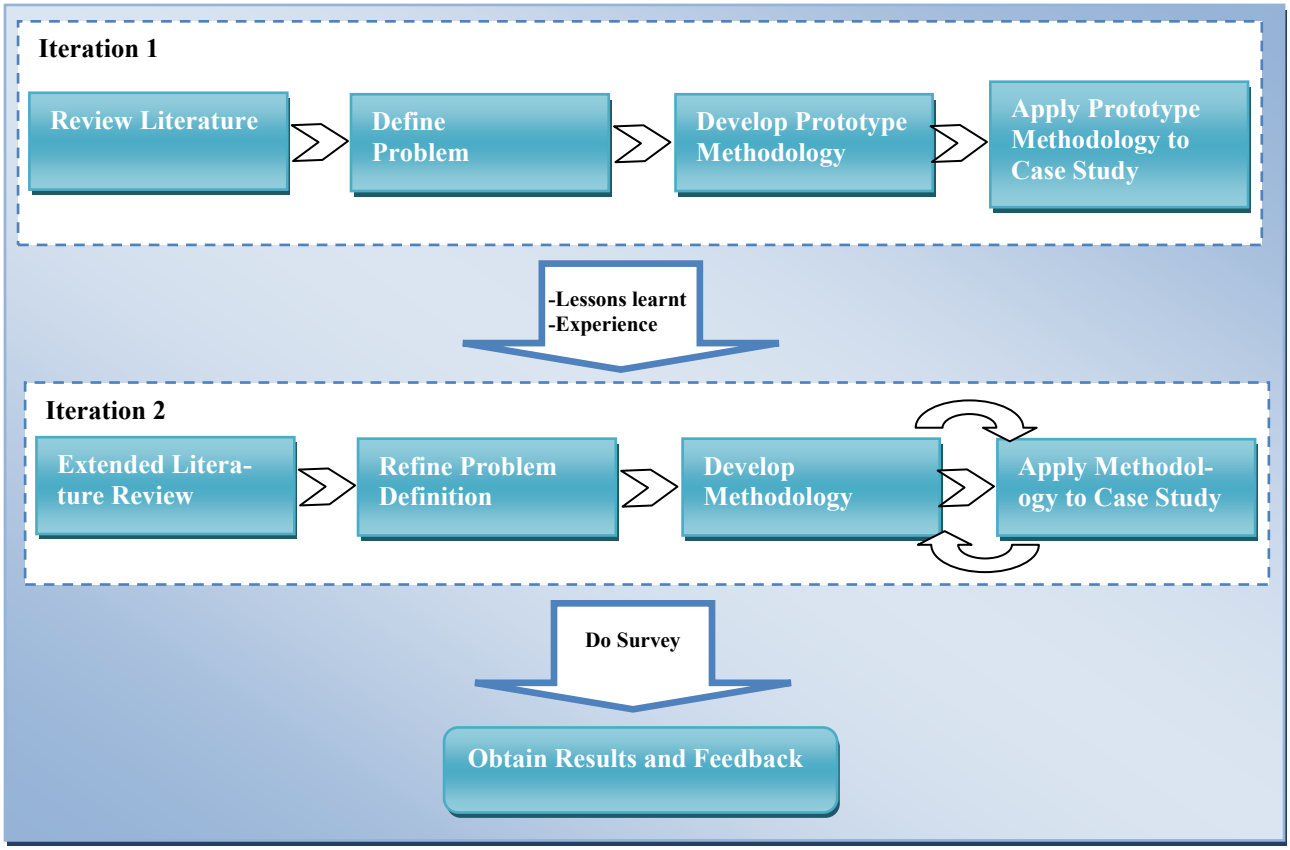


Figure 1. Research Method

The first step of the research method consists in a preliminary literature review. This step permitted to gain an understanding of the researched topic and therefore to gather and analyze its main concepts, ideas, and challenges. Initially, scientific material was reviewed to understand the current methodologies used within the health care domain for the development of metrics, as well as their reported difficulties. The research was not only restricted to the health care domain but was carried throughout other industries as well. In this way the investigation could benefit from different ideas than those found in the health care context. For example, methodologies used for the

discovery and generation of metrics within the software development sector were also investigated.

After this step, the researcher had collected enough information to define specific issues within the explored area. The main problem was stated as "the need to find a way of addressing the existing gap between the data collected and the development of metrics that can report on this data while addressing some of the shortcomings outlined in the literature". Having a clear objective initiated the process of a search for a solution.

A *prototype methodology* (i.e, an initial, tentative version) was put together by using key concepts of an approach originally created for the software development industry: GQM (Goal Question Metric) [Basili, 2005]. Besides the creation of metrics, the prototype methodology also laid out the need for developing dimensional models to support the reporting of metrics.

The next step was to test the solution with a case study. The implementation of the prototype methodology on a real scenario (related to adverse events) helped determine its valid steps as well as the difficulties encountered during the test exercise. In other words, the gathered experience provided valuable insights on how the proposed prototype had performed. This step led to a realistic approach to what was needed to develop a methodology for the discovery of metrics in the health care context.

With the obtained results, the previously described steps took place again. A deeper literature review was performed to address the difficulties encountered during the implementation of the prototype methodology. The problem was refined and a more robust methodology was drawn. This solution was then tested on a different case study (incident reporting), providing the research with a "double loop" validation process. It should be mentioned that while using the methodology, feedback was gathered, and the methodology was improved and then re-applied. Finally, the results were collected and presented.

A survey permitted to collect the opinions of the health care stakeholders involved in this study, and to assess the suitability and impact of the methodology.

1.7. Thesis Contributions

The contributions of this thesis are:

- The creation of a goal-driven methodology, called *Health Care Goal Question Metric* (HC-GQM), for the development, collection, and analysis of meaningful metrics in the context of health care quality.
- The adaptation of the Goal Question Metric (GQM) approach, a measurement methodology created for the software development industry, to the health care domain.
- The application of the methodology to two health care quality contexts: an Adverse Event Management System (in iteration one) and an incident reporting system called Patient Safety Learning System (in iteration two) at a teaching hospital.
- The elaboration of "Patient Safety" reports that meet the needs of real health care stakeholders involved in the measurement exercise.
- The creation of reports that contain metrics enabling the business intelligence of health care data.
- The provision of a road map for health care institutions that might want to apply the goal-based methodology in order to obtain tailored metrics and reports that are aligned with their goals.

1.8. Thesis Organization

Chapter 1 has presented an introduction to the research and its application domain. Chapter 2 contains information regarding the literature review process. It describes analyses and compares existing methodologies, technologies, and tools that are used to address the development of metrics, the modeling of dimensional data, and the creation of BI reports. Chapter 3 addresses the evolution and implementation of a first prototype methodology to solve the research problem. It also describes the experiences gathered from its implementation on an Adverse Event Management System at a teaching hospital. Chapter 4 elaborates on the assembly of a refined HC-GQM methodology capable of addressing the thesis research problem while taking into consideration the difficulties arising from the first case study. Chapter 5 describes the implementation of the methodology in the context of the Patient Safety Learning System at the same hospital, together with the results of a survey on the usefulness of HC-GQM. Chapter 6 discusses the threats to the validity of this work together with the conclusions and future work.

Chapter 2. Metrics Development Methodologies and Business Intelligence

This chapter aims to review and critically analyze literature that can support and guide the development of this research. The following three topics have a significant impact on the investigation been carried out: performance and quality metric development methodologies, measurement methods in health care, and business intelligence.

The key objective of this literature review is to find recognized methodologies and frameworks that address these three topics. It will mainly focus on their goals, steps, success factors, and drawbacks. Furthermore, comparative analyses will be performed and the most suitable methodologies will be used towards the investigation.

A secondary objective is to briefly introduce business intelligence tooling, with a particular emphasis on IBM Cognos 8.

2.1. Software-oriented Metrics Development Methodologies

Metrics, as a measurement tool, provide a means of assessing the state and evolution of any institution, entity or process. However, the identification of metrics it is not a simple task since too many metrics are generally identified [Phelps, 2004]. This problem tends to overwhelm stakeholders involved in their interpretation and analysis. Therefore, it becomes necessary to investigate methods of identifying the minimum number of metrics that stakeholders need to assess their institutions, entities or processes.

The software industry has created many frameworks that address this issue. Examples of those are: the Goal Question Metric (GQM) approach [Basili, 2005], Personal Measurement Software (PSM) [Card, 2003-a p738], and the framework developed by [Fenton and Pfleeger, 1997] in *Software Metrics a Rigorous and Practical Approach.* The following subsections review each of these methodologies.

2.1.1 Goal Question Metric Approach

The Goal Question Metric (GQM) approach was created by Victor Basili and his colleagues in the 1980s, in partnership with the NASA Software Engineering Laboratory [Shull et al., 2006]. This methodology focuses on the development, collection, and analysis of a set of variables (or indicators) that are capable of addressing the progress of defined measurement goals in a quantifiable way.

The core of this approach is based on a top-down hierarchical structure formed by three levels: conceptual, operational, and quantitative. At the conceptual level, a set of measurement goals are developed. They take into account: a measurable object, quality models, different points of view and the object's context. The operational level expands on the goals by developing a set of questions that describe the object with respect to a quality or performance problem. The quantitative level associates quantifiable variables to the determined questions. These variables can be either objective or subjective, depending on what they are measuring. Within this structure, some metrics can be used to answer different questions under the same model, and different models can have some questions and metrics in common.

This approach can be described as a 6-step framework: [Basili, 1993; Basili, 2005]

1. Create a group of business goals and their related measurement goals.
2. Pose questions capable of characterizing in a quantifiable way the developed goals.
3. Find the metrics that can better provide an answer to the questions.
4. Develop instruments to collect the data.
5. Gather, validate, and analyze the collected information to take corrective action.
6. Analyze data in order to see whether it addressed the goals or not.

It should be pointed out that after analysing the papers from Basili [1993; 2005] it was discovered that despite the many years in use and the radical changes in technology, this framework has not experienced dramatic modifications since its creation.

Some authors suggest that the implementation of the GQM approach does not have to be as sequential as stated by Basili [Van Solingen et al., 1999]. Instead, the 6 steps' inputs and outputs should vary depending on the context of the implementation and the scope of the project. Many of the steps can intermingle and have dependency relationships with the processes that implement them.

GQM can be considered as an open methodology that can be utilized "whenever effective metrics are needed to assess satisfaction of goals" [Khamis et al., 2007]. This means that the approach is applicable to a diverse range of organizations, environments, products, processes or resources. Nowadays, GQM is already considered a "de facto standard for the definition of measurement frameworks" [Berender et al., 2006 p.316].

The main benefit of implementing GQM is the achievement of measurement goals [Van Solingen and Berghout, 1999]. However there are many other positive impacts of this methodology such as "improving communication within a project team, attitude of personnel, process definition and process execution" [Van Solingen and Berghout, 1999]. By improving the definition and execution of processes, risks can be mitigated and quality increased. For these reasons, GQM could be also implemented as part of a quality improvement process.

Weaknesses have been reported for the GQM approach as well. The most outstanding issue is the risk of identifying more metrics than are possible to collect or analyze [Berender et al., 2006 p316]. This could also translate into "a top-down approach [that] ignores what is possible to measure at the bottom." [Bache & Neil, 1995]

In order to solve this problem, an extension of the GQM approach is proposed by Berender et al., [2006] where "prioritization" tools are used for limiting the number of metrics identified and "categorization" tools are used for the balancing of different dimensions.

These extensions make the approach more in-line with the ISO/IEC 15939:2002 standard (reviewed in the next section), which mentions that in order to make a measurement framework efficient, a smaller group of the recognized metrics should be gathered [Berender et al., 2006 p316].

The "improved" GQM approach [Berender et al., 2006] is very similar to GQM. It provides a set of best practices for developing goals and questions and it uses categorization and prioritization techniques before developing the metrics. The former process classifies questions according to their characteristics in such a way that the resulting questions cover several dimensions instead of only a few. Sometimes, the output of this categorization process can be used to determine if more questions need to be formulated. The latter handles the prioritization of the questions and goals developed. Since their relevance, urgency or importance generally differ, it is wise to develop the metrics for the more important or urgent ones. Many prioritization techniques can be used for this purpose [Berender et al., 2006 p318].

Another reported issue is related to the methodology's lack of iterative behaviour [Kurian, 2009]. GQM, as stated in [Basili, 1993], does not explicitly address its steps in an iterative way. This issue brings along problems of effectiveness, since due to stakeholders' inexperience with the methodology they might not focus enough on the metrics to be developed [Kurian, 2009]. Further, with a non-iterative approach, stakeholders are forced to think about and express all of their needs at once. This means that after finalizing the study if certain goals or questions were not considered they would not be included in the measurement exercise. As an example, constraints related to privacy and disclosure of information could be imposed after the fact. Without an iterative process, these issues could not be addressed in a timely way. Therefore, reports disclosing "confidential" information could be prevented from being used.

Studies have been performed to address this difficulty. *Continuous GQM* and *Validating GQM (V-GQM)* [Kurian, 2009] are two examples of such studies.

2.1.2 Practical Software and Systems Measurement (PSM)

PSM (Practical Software and Systems Measurement) is a systematic information-driven approach to software measurement for project management purposes [Card, 2003-a p738]. It was established in 1994 and sponsored by the US Department of Defence and the US Army [PSM, 2010].

This methodology has become widely accepted in the industry as the measurement approach for the management of software-intensive system development projects [Card, 2003 p739] and it has been adopted by the US government and industrial organizations. Furthermore, in 2001, the PSM process was published as the international standard ISO/IEC 15939:2007 [ISO, 2007].

The approach has been developed to address project managers' needs, specifically at the project level [Card, 2003-a p738; Card, 2003 p2; PSM, 2010]. The main idea behind PSM is to provide organizations with the necessary tools that enable the approach to be tailored to their specific information needs (or goals). Some of the highlights of this methodology are its focus on costs, schedules, and technical objectives, as well as the identification of project-specific issues [Gold Practices - DACS, 2010].

According to PSM Methods of Operation [PSM, 2006], the objectives of this methodology are:

- To establish a proven process to implement a tailored information-driven measurement process for software and systems engineering management.
- To provide a basis for objective communication and informed decision-making.
- To establish a foundation for organizational and executive-level performance management.

PSM is based on a combination of the Measurement Information Model and the Measurement Process Model [McGarry et al., 2002]. The former model is in charge of the structure of the measurement project, meaning that it provides connection between the information needs and the entities to be measured [McGarry et al., 2002]. Furthermore, the Measurement Information Model provides a well-defined analysis path to improvements suggested by the collected data [McGarry et al., 2002 p.1-12]. The Measurement Process Model is in charge of providing a guide to the measurement activities and tasks to be executed [McGarry et al., 2002]. This model is of special interest to this research since it outlines the steps that are taken in order to discover and develop the metrics. Specifically, this model encompasses four activities: Establish and Sustain Commitment, Plan Measurement, Perform Measurement, and Evaluate Measurement [Card, 2003]. As showed in Figure 2, Plan Measurement and Perform Measurement are considered the two core measurement activities.

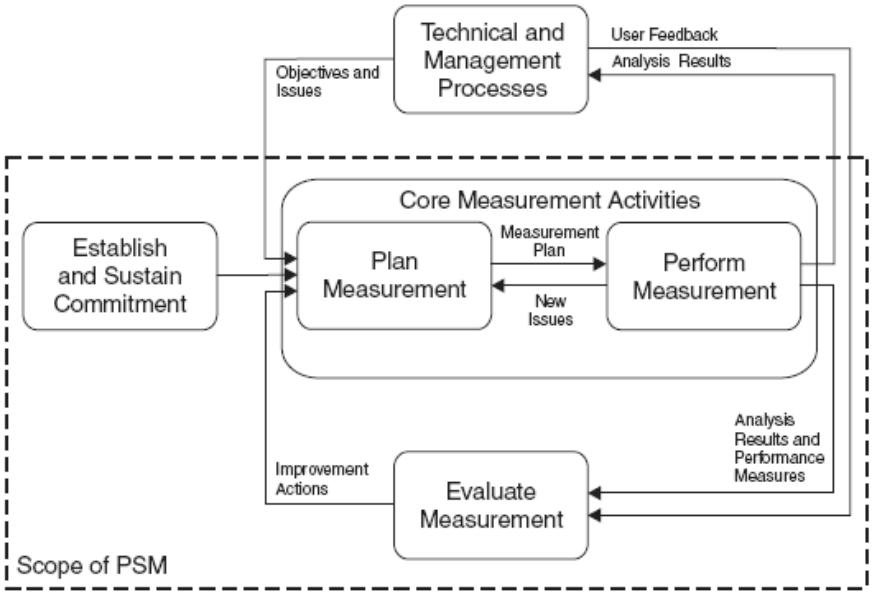


Figure 2. Measurement Process Model [McGarry et al., 2002]

The *Establish and Sustain Commitment* activity ensures the required support from management at the various levels [McGarry et al., 2002]. It also establishes the resources, tools, and training needed to develop and successfully implement the measuring effort [PSM, 2004].

The *Plan Measurement* activity identifies decision-makers project-specific issues [PSM, 2004]. Then it aids with the selection of the measures needed in order to address the identified needs using the "Measurement Information Model" [McGarry et al., 2002]. This activity also encompasses the "definition of data collection, analysis, and reporting procedures" [McGarry et al., 2002]. According to PSM's Measures for Success [PSM, 2004], the measurement plan should start by implementing a small set of metrics and incrementally add additional metrics. The plan should clearly state "What" will be measured and "How" the process will work.

The *Perform Measurement* activity performs data collection, performance analysis, and the presentation of results [McGarry et al., 2002; PSM, 2004]. Experience implementing this activity [PSM, 2004] suggests that the "data collection" should be automated whenever possible; the "analysis" should include estimation, feasibility analysis of plan, and performance analysis of the actual data against the planned values and the "reporting mechanisms" should be identified at the various levels of the organization, but mainly at the project level, since project managers are the targeted users of this methodology.

The *Evaluate Measurement* activity states that the measurement process and the measures should be periodically revised and evaluated [McGarry et al., 2002]. This step closes a loop in the framework converting the process into an iterative effort: the organization implements improvement actions, "information needs" change, and therefore the metrics are refined.

Although "PSM and ISO/IEC 15939 provide a common language that may improve communication on software measurement issues between practitioners and researchers" [Card, 2003-a p739], they also present some shortcomings that are associated with the PSM framework itself. For example: "PSM does not define all measurement procedures such as those needed to address project-specific information needs, different software domains or individual system technologies" [PSM, 2006]. Also, PSM focuses on using existing metrics rather than creating new ones and it does not involve all stakeholders of an organization.

2.1.3 Software Metrics—A Rigorous & Practical Approach

"Software Metrics—A Rigorous & Practical Approach" is a measurement framework developed by Fenton and Pfleeger for the software industry. Two editions of a book with the same name [1991 and 1997], as well as some published articles, explain the concepts, steps, and rationale of this software measurement framework. According to Oman and Pfleeger in their book "Applying software metrics" [1997], the first version of the framework shows how measurement requires the definition of entities and their attributes as well as the relation of these attributes to values, units, and scale types [Oman & Pfleeger, 1997]. Furthermore, this initial framework elaborates to establish a measurement basis for software metrics activities [Fenton & Neil, 1999 p.2].

The evolution of the framework continued as definitions of measurement goals were introduced to drive the measurement exercise [Fenton, 1994 p200]. To accomplish this step, an improved version of Basili's Goal-Question-Metric approach was developed [Fenton & Pfleeger, 1997]. In addition, a process maturity framework capable of providing information about the availability of metrics depending on the project's state was also added [Fenton & Pfleeger, 1997 p88]. This means that only the available metrics would be developed, tackling one of GQM's reported issues: the definition of too many metrics.

The process maturity framework elaborates on the structure of the Capability Maturity Model (CMM) created by the Software Engineering Institute and it takes into account not only the maturity of the process but also the maturity of its outcomes [Curley, 2006 p.1]. According to the CMM, there are five levels of process maturity: initial, repeatable, defined, managed, and optimized [Ho, 2010].

The result of many years of improvement is a methodology based on three principles:

- Entity classification;
- Creation of relevant measurement goals;
- Determination of the level of Maturity of the project.

These principles can be appreciated in the approach's steps, which are laid out in [Fenton & Pfleeger, 1997]:

1. Identify a set of *entities* to be measured and their attributes. Entities are the objects to be measured. They can be grouped as processes, products or resources [Fenton & Martin, 2000 p.5; Fenton, 1994 p199]. Each entity has a set of measurable attributes (internal or

external) which are essential in that they provide the data for the metrics. Internal attributes measure inner aspects of the entity separating it from its behaviour. External ones measure how the entity interacts and relates to its environment [Fenton & Pfleeger, 1997; Fenton, 1994].

2. State a set of measurement goals. Step one of Basili's GQM approach is introduced to define a set of measurement goals to guide the exercise [Fenton & Pfleeger, 1997].
3. Pose questions that characterize goals in a quantifiable way. Step two of Basili's methodology is followed to characterize in a qualitative way each one of the stated goals [Fenton & Pfleeger, 1997].
4. Generate metrics to answer the questions. This is obtained by measuring one or more attributes of one or more entities. When no attributes of entities are available, then objective metrics cannot be collected. Subsequently, subjective metrics should be introduced [Fenton & Pfleeger, 1997]. At this point, it is not possible to tell whether the project is being objective and feasible and it can actually provide the necessary information for the measurements [Fenton & Pfleege, 1997 p87].
5. To resolve this, the "process maturity" framework is taken into account [Fenton & Pfleeger, 1997]. The level of maturity exposes what metrics are available at a certain point in time. This definition provides certain boundaries for the metrics collection [Fenton & Pfleege, 1997 p88]. According to the selected level of maturity, it is possible to determine which metrics can be collected in order to answer the questions that define a goal.
6. Validate the metrics. Each metric represents the value of an empirical attribute. The validation process ensures that logic is maintained within empirical values. The validation that is used depends on the type of variable analyzed. This occurs on a case by case scenario [Fenton & Pfleeger, 1997].

By combining GQM with CMM, this methodology addresses some of GQM's reported issues such as the extensible number of metrics created. However, this combination can be regarded as a drawback as well. According to Bach [1994], CMM focuses on processes but ignores people and displaces goals of improving processes to acquiring a higher maturity level. These issues counteract the purpose of developing goals and making use of peoples' (stakeholders) criteria, two of the basics of GQM.

2.1.4 Methodologies Comparison

Table 1 is a reference table with a summary of the three methodologies discussed above. This table is meant as a quick reference point to compare and contrast the key points.

	GQM	PSM	Software metrics. A rigorous and practical approach
Goals	- to develop, collect, and analyze of a set of variables (or indicators) that are capable of addressing in a quantifiable way the progress of defined measurement goals	- to implement a measurement process for software engineering management - to provide a basis for objective communication and informed decision-making - to establish a foundation for organizational and executive-level performance management	- to establish a measurement basis for software metrics activities - to combine the definition of entities with GQM and process maturity to form a more robust measurement methodology
Steps	- create business and associated measurement goals - generate questions - specify measures to answer the questions - develop methods for data collection. - collect, validate, and analyze data in real time - analyze data in post-mortem fashion	- establish and sustain commitment - plan measurement - perform measurement - evaluate measurement	- identify a set of Entities and their attributes - state measurement goals - pose questions to define goals - generate metrics to answer the questions. - map project's state to a level of maturity - validate the metrics
Success factors	- variable scope of implementation - adapts to different organizations and environments - is applicable to all life-cycle products, processes or resources - is considered a de facto standard for the definition of measurement frameworks	- was published as the international standard ISO/IEC 15939:2007 - addresses project manager's needs, specifically at the project's level - focus on costs, schedule, and technical objectives - the identification of project specific issues	- uses GQM to define goals that guide the measuring. - addresses issues related with GQM by using process maturity framework.
Shortcomings	- identification of more metrics than are possible to collect or analyze - does not address each step in an iterative or incremental manner	- does not define procedures to address different software domains or individual system technologies - focuses on using existing metrics rather than creating new ones	- focuses on processes but ignores people - displaces goals of improving processes to acquire a higher maturity level

Table 1 Methodologies Comparison

2.2. Metrics Development Methodologies in Health Care

Over the past years many approaches have been developed to measure performance and quality of care in hospitals. Many of these approaches focus on developing a "standard" set of metrics with the objective of using them as "benchmarking" tools in hospitals. This is the case of "The Specifications Manual for National Hospital Inpatient Quality Measures" developed by the Joint Commission [The Joint Commission, 2010]. Other methodologies suggest the use of pre-existing guidelines for "high quality" care and integrating them with the measurement frameworks. The "Methodology for the selection and creation of performance measures for quantifying the quality of cardiovascular care" developed by The American College of Cardiology and American Heart Association [Spertus et al., 2005] is an example of such approach. Its main objective is to measure the adherence to clinical practice guidelines as a way of quantifying quality of care.

Yet other methodologies are developed with the purpose of obtaining indicators capable of providing information that can be publicly reported upon. An instance of such methodology is outlined in "Strategy-based system-level cancer care performance indicators", an approach developed in Ontario as part of an initiative to improve the quality of care in cancer-care organizations [Greenberg et al., 2005]. The indicators developed are not a measure of hospitals' internal evaluation but rather a way of demonstrating that the health system was accountable and monitored as a whole [Greenberg et al., 2005].

This chapter focuses on methodologies developed by and for the health care sector, whose main purpose is to discover and create indicators to measure quality. Section 2.1.1 reviews the IQIP approach [IQIP, 2010] while Section 2.2.2 outlines the PATH experience [Veillard et al., 2005].

2.2.1 IQIP

The International Quality Improvement Program (IQIP) was established in 1985 in Maryland, USA, as part of a research project conducted by 7 hospitals and the Maryland Hospital Association [Kazandjian et al., 1995 p39]. By 1994, over 900 hospitals across USA, 5 hospitals in England and one in Japan had joined this initiative [Kazandjian et al., 1995 p39]. Today, this methodology "serves the performance measurement and safety improvement needs of health care organizations worldwide" [IQIP, 2010]. It is also considered the biggest data repository of quality indicators [Thomson et al., 2004 p51].

The main objective of this approach is the provision of defined sets of indicators capable of providing insights on quality and performance. IQIP engages in activities like searching for "the most valid indicators, the most reliable methods of data gathering and the optimal clarity of analysis presentation" [Kazandjian et al., 1995 p40], which makes the reliability and relevance of the proposed indicators quite high. Indicators were developed using a combination of a strong literature review, peer reviews with expert panels as well as through the implementation of pilot projects [Thomson et al., 2004 p52]. All of these methods are fairly known and used by other methodologies. However, what makes IQIP different is the support and training provided to users, like hospitals' coordinators, on how to use, evaluate, and understand the outcomes of the indicators [Thomson et al., 2004 p52; Kazandjian et al., 1995 p42].

These indicators are designed to tackle performance "through statistical and epidemiological techniques in a value-free manner." [Kazandjian et al., 2005 p162] This means that the values of the indicators by themselves do not indicate whether a performance is good or bad, they are merely flags to direct attention to certain areas. For example, if rates of death are compared between 2 institutions or departments (X= 15% and Y=10%), the fact that X has a higher value than Y does not necessarily mean that X's performance is worse than Y. The disassociation of values that results is further explained by Kazandjian, who states that even when many indicators imply "safety" their measures of safety or performance is not as accurate when used alone [Kazandjian et al., 2005 p163].

IQIP, as a methodology, follows three main principles: "measurement of quality", "indicator reliability and validity" and "usefulness of institutional trends, patterns, and profiles" [Kazandjian et al., 1995 p43].

Measurement of quality describes how indicators do not measure quality of care on their own, but the quality of the collected data. What really measures quality is the people involved in understanding the indicators, which serve the purpose of pointers to certain "possible" problems [Kazandjian et al., 1995].

The indicator reliability and validity principle exposes how indicators should be valid in order to be reliable, and their performance varies depending on who collects them [Kazandjian et al., 1995].

The third principle questions how probable it is to obtain trends and patterns by measuring and monitoring the rates of certain institutional indicators. Furthermore, it questions if these

trends and patterns can be used towards comparing performance among different hospitals [Kazandjian et al., 1995].

Some of the key aspects of success of the methodology are reported in [Thomson et al., 2004]:

- Voluntary hospital participation in the implementation of this approach.
- Anonymous feedback system that supports quality improvement activities.
- Lack of publishing requirements of the results of the study to external parties.
- A user-driver system. This means that the project takes into consideration users' viewpoints to drive the selection of certain sets of indicators to be implemented. This is obtained through the use of surveys and consultations.
- And importantly, the support and training offered to assist in the collection, interpretation, and further use of the indicators.

These key success factors might explain the worldwide implementation of the methodology. Figure 3 shows all the countries, alongside the number of hospitals, where IQIP has been implemented [IQIP, 2010].

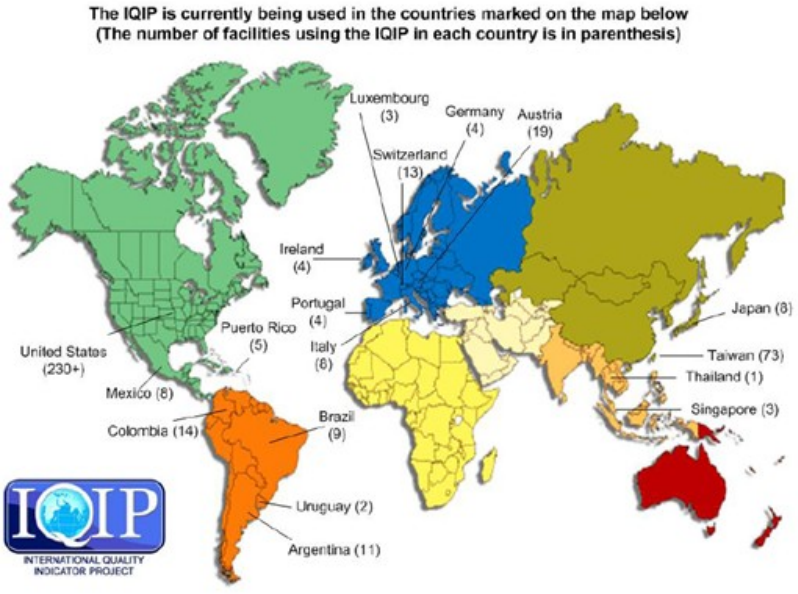


Figure 3. IQIP Implementation Worldwide

Some issues related to IQIP are reported in [Kazandjian et al., 1995]:

- Some of the discovered indicators using this methodology had imperfections. This did not mean that they were not valuable, but users should be aware of their limitations when using them.
- The project had restrictions regarding the field of health care research. This means that it was unclear that a determined action would yield a specific result (i.e., the relationship between a process and its outcome was weak and unpredictable).

2.2.2 PATH

PATH stands for Performance Assessment Tool for quality improvement in Hospitals [Groene at al., 2008]. Its main objective is the provision of a tool to help hospitals assess their performance, question their results and act upon them. In this way, PATH also contributes towards improving the quality of care inside the hospitals [Veillard et al., 2005].

PATH does not propose a methodology to support hospitals, or other health care centers, to develop their own indicators. Instead, it suggests a path for the implementation, collection, and analysis of a group of 18 "already-developed" indicators that measure performance in 6 different, yet interconnected dimensions: clinical effectiveness, efficiency, staff orientation, responsive governance, safety, and patient centeredness [Groene et al., 2008].

Although PATH does not provide guidelines on how to develop metrics, literature such as [Veillard et al., 2005] does outline the methodology followed to develop the indicators that hospitals utilize to assess their performance.

The development of PATH was created in three stages: the creation of framework containing indicators capable of measuring performance, the implementation of this framework in 8 different countries, and lastly the development of guidelines to facilitate the benchmarking of the framework [Veillard et al., 2005]. The first stage is further analyzed by this research since it provides insightful information for the investigation carried.

PATH started by forming a team of hospital performance experts from all over the world. They participated in workshops where literature was reviewed, and their individual expertise on performance measurement for the health care was shared [Veillard et al., 2005]. As part of the development of the framework a detailed analysis of existent methods and models of performance was studied [Veillard et al., 2005].

The PATH conceptual model was developed as a result of these activities. It was based on the six inter-related dimensions mentioned above. These are the attributes that best described the concepts and models of performance [Veillard et al., 2005 p2; Groene et al., 2008 p156].

Armed with the knowledge acquired through the literature review, the workshops as well as surveys, an initial list of indicators was drawn and then iteratively reduced [Veillard et al., 2005]. This reduction was made possible through a mechanism where indicators were ranked according to their importance, relevance, usefulness, validity, and reliability [Veillard et al., 2005]. Supporting evidence for certain indicators led to discoveries that guided the refinement of the conceptual model. The result of these steps was a set of indicators with their "rationale, operational definition, data collection issues, [and] support for interpretation" [Veillard et al., 2005 p5].

After refining the list, an operational model was created. This model pointed out how indicators related to each other and to quality improvement dimensions formed by the conceptual model.

The methodology used to obtain the indicators and the performance dimensions can be summarized as follows:

- The construction of a conceptual model based on dimensions of performance.
- The gathering of a first list of indicators based on literature reviewed.
- The development of an operational model to reduce the initial list of indicators to only those that are sufficient and capable to measure the dimensions of the conceptual model.

Groene et al., [2008] have reported some positive outcomes derived from the implementation of this project. It can be remarked that PATH made it easier to bring together intrinsically different quality assessment activities such as "quality improvement in different departments" [Groene et al., 2008]. Further, in many countries like Belgium or Slovakia, PATH served as a trigger to the initiation of other indicator development projects [Groene et al., 2008].

After the implementation of PATH in several countries, Groene et al., [2008] have reported some issues. Among those are:

- The framework intends to compare data from several hospitals across several health care systems. This proves difficult as different systems have different expectations and characteristics, as well as hard-to-reconcile data sources.

- PATH implementation is not easy when hospitals already have other performance measurement methodologies in place because it causes competition and it is hard to find a niche for the project.
- The project requires specialized personnel (i.e., data capturers) that hospitals may not have in sufficient numbers.
- Lack of time to collect data, lack of resources and funding, as well as organizational inertia also hinder the implementation of the approach.
- Clashing of indicator definitions that are already collected by other frameworks.
- Some indicators had to be adapted depending on the context, resulting in vague indicators that are hard to operationalize.
- Not all reports were successfully understood by stakeholders.

To conclude, despite the fact that PATH is not a "metrics development methodology", the experiences gathered from the literature provide an insight on steps that should be taken into account when discovering metrics as well as some of the particularities of the health care sector when it comes to implementing a measurement project.

2.3. Business Intelligence Tools

2.3.1 Overview

BI systems generally store data in data warehouses which are used as the foundation of a wide variety of analyses such as: simple reports, slice and dice, drill down, ad-hoc queries, forecasting, real time analysis etc [Ballard et al., 2006]. The outcome of these analyses is the provision and delivery of actionable information in the right format, to the right decision-makers whenever they need it.

In order to obtain the aforementioned analysis's types, analytical tools are needed to process the data stored in data warehouses [Ballard et al., 2006; Lane et al., 2002]. The simpler the analytical tool the simpler the question that addresses. For example, ad-hoc reports, queries, and drill down reports help answering questions of the type: what, when, how much, and where. More complex tools such as statistical analysis, forecasting, extrapolation, and optimization help

with the "why" aspects [Negash et al., 2008 p180]. Analytical capabilities need to have in place suitable people and organizational infrastructures [Negash et al., 2008].

According to IBM [2009], reporting is a key element of performance management since this helps accessing outstanding information that managers use to take action. At the same time, reporting is also a complicated element due to the wide variety of users and roles that organizations hold and that have different information needs.

A recognized issue with current reporting tools is that it is nearly impossible for users to create new reports or modify those they currently use [IBM, 2009 p15]. It is important that stakeholders across the organization be capable of accessing data in a fast and user-friendly fashion. Reporting software should permit collaboration and the sharing of information across team works, thereby increasing the productivity and efficiency of processes and leading to cost savings [IBM, 2009 p15].

Many BI tools are available in the market, including IBM Cognos 8 [IBM, 2009], SQL Server Reporting Services [Microsoft, 2010], Oracle Business Intelligence Suite [Oracle, 2010], SAP Business Information Warehouse [SAP, 2010], SAS Business Intelligence [SAS, 2010], and HP Neoview [HP, 2010]. In the next section, IBM Cognos 8 is reviewed in more detail as it is the BI reporting tool that was used by TOH and the University of Ottawa in their adverse event detection project.

2.3.2 IBM Cognos 8 BI Tool

IBM Cognos 8 BI is a tool based on web services that provides many useful capabilities besides basic reporting features [IBM, 2009]. Examples of those are: analysis, score-carding, event management, inventory lists, and high impact dashboards, to mention a few [McMillian, 2007]. This tool can integrate into an organization's IT infrastructure without having to include additional storage space or redundant environments [IBM, 2009].

Reports created with this tool can contain different sorts of objects such as charts, tables, lists, images, logos, and even embedded applications that can be linked to the reported information [IBM, 2009]. Data presented in reports do not have to be necessarily extracted from one particular data source [IBM, 2009]. This means that a single report can query diverse sources and information is still brought together in context.

According to IBM, Cognos 8 BI is a product that delivers a "complete range of BI capabilities in a single proven architecture" [IBM, 2009 p4]. In other words, this architecture is characterized by:

- Multi-tier architecture with open web standards such as XML, SOAP, WSDL [Mouchakkaa and Danielewski, 2006; IBM, 2009].
- Application Programming Interfaces that let programmers customize capabilities using various languages including .Net, C, and C++ [IBM, 2009].
- Peer-to-peer multithread architecture that provides a distributed environment [IBM, 2009].
- A complete scalable model, as it can go from a handful of users to hundreds of thousands [http://www.cognos-bi.info/cognos8.html; IBM, 2009].

IBM Cognos 8 also addresses three important reporting requirements:

- Recognition of different user types [IBM, 2009]: Reports, depending on the user's role, will know what information to present and the specific format needed. This functionality is useful since each work group will have the information they need, and will not be overwhelmed by extra capabilities. This also enables proper access control to the information.
- Adaptation to multiple data sources [IBM, 2009]: The tool gives access to reports to collect information across heterogeneous data sources. For example, a report can be sourced by multiple data warehouses and XML files at the same time. Database connections can be dynamically set based on session parameters. The main goal of this multiple data sourcing is to give users a complete view of a specific business issue that can crosscut operational environments.
- Support for many report types [IBM, 2009]: This tool supports a wide range of reports from ad hoc queries and managed, business, dashboards reports (which require high usability and interactivity) to invoices, statements, and bills (which require sophisticated formatting and high scalability). Since it is browser based, Cognos 8 does not require the installation of specific desktop applications, which simplifies deployment and maintenance.

In addition, reports created with Cognos BI tools can support multiple languages. Reports (including both user's interface and content) will adapt to the end-user's language [IBM, 2009].

2.4. Conclusion

This chapter presented well-established methodologies, approaches, and tools that were used in the development and implementation of metrics in the software industry as well as in the health care sector. Each section discussed the objectives, main steps, success factors, as well as drawbacks of a particular methodology. We can observe that goal-driven approaches successfully used in the software industry (e.g., GQM) are not really used in health care, and that health care has specific needs and constraints not usually taken into account by existing goal-driven approaches. Also, a brief review of business intelligence was conducted, with a particular emphasis on the capabilities of IBM's Cognos 8 tool for report generation.

The next chapter will describe the design and prototype implementation of a new health-care-oriented, goal-driven methodology for the development of metrics and reports.

Chapter 3. Prototyping of a Goal-Driven Methodology

The current chapter explains the need for the development a specific goal-driven methodology to provide a teaching hospital in Ottawa with meaningful metrics to report on, together with the details of this methodology. It also discusses the success factors and issues encountered during its first implementation as a prototype.

Section 3.1 provides background information on an Adverse Events Management System developed for The Ottawa Hospital (TOH). Section 3.2 presents why GQM is selected as the foundation of the new goal-based methodology and then explains the approach. Section 3.3 outlines the implementation of the prototype methodology at TOH, whereas Section 3.4 discusses the difficulties encountered during this implementation.

3.1. Adverse Events Management System

The Adverse Events Management System (AEMS) is an application developed for the collection and analysis of certain events that are related to hospitals' inpatients during their period of care. It was created by a group of software engineering students from the University of Ottawa in 2009 [Blais et al. 2009] following the specifications and business requirements laid out by epidemiology researchers at the Ottawa Hospital Research Institute (OHRI). The purpose of this application is to collect and track AE-related data for pilot implementations of an AE detection process throughout different clinical units of The Ottawa Hospital. AEMS is based in part on a previous prototype by Behnam et al. [2009], but with a larger scope, a more usable interface, and a more robust, secure, and extensible implementation.

To implement such application, it was necessary to take into consideration the AE detection method that the system would model. AE detection methods generally focus on their ability to discover the greatest number of AEs within health-related organizations as well as the sources of data to be used for this purpose. Generally, these methods make use of certain common

"screening criteria" that help them define the occurrence of AEs. Walsh presents an example of such list [Walsh, 2000 p48]. Some of its values include:

- Antibiotic or drug use problems;
- Abnormal laboratory, radiograph or other test results not addressed by physicians;
- Hospital acquired infections;
- Cardiac or respiratory arrest;
- Transfers from general care unit to special care unit.

The core idea behind identification of AEs is to analyze one or various sources of data on AEs (such as chart reviews or medical record reviews) and try to match their content with the list of *triggers* that are usually symptoms or causes of AEs. The occurrence of any of these triggers raises flags that specialists can carefully follow.

Among the most recognized approaches are one-stage and two-stage chart reviews, medical record reviews, self reporting (usually done by physicians and nurses), administrative data review, clinical surveillance and patient care observation [Thomas and Petersen, 2003]. All of them have strengths and shortcomings that somehow make their use a better or worse option depending on the organization and its available data. Researchers at OHRI decided to investigate the use of prospective adverse event surveillance, a form of clinical surveillance, as their AE detection method.

In this prospective surveillance process, an observer nurse inputs specific data related to certain triggers into the AEMS, on a daily basis. The information collected is then revised by a committee of doctors (the reviewers) on a weekly basis. This committee is necessary because medical professionals generally base their analysis on their previous experience and skills [Walsh, 2000]. The reviewers analyze the figures gathered by the nurse and decide together whether a particular event is adverse or not [Behnam et al., 2009]. In case of a positive identification of an AE, this tool allows reviewers to classify the type of error committed as well as the degree of harm caused to the patient. The system also gathers relevant incident data such as the event date and time of occurrence, its location, and the trigger that originated its capture. Patient information, including age and chronic illnesses, is also collected.

AEMS is a Web-based application that follows a three-tier architecture composed of a "Web browser on a tablet PC connected through a wireless network, a Web server containing a

presentation layer and a business logic layer, and a database server containing patient information and stored procedures" [Behnam et al., 2009 p9]. The system was also developed in such a way that different levels of access and tasks were granted to the users depending on their role within the organization. For example, observer nurses are granted with privileges for creating and documenting event while reviewers (doctors and researchers) use the application to categorize events and decide on their adverse nature and impact, if any.

3.2. A Goal-driven Methodology Prototype

3.2.1 Situation Preceding the Prototype

AE detection methods are good for collecting data that supports the discovery of AEs. However, if adequate reporting mechanisms are not present, researchers and other interested stakeholders cannot extract any meaning or *information* out of the collected data.

Several pilot projects were conducted with AEMS within different clinical units of the Civic campus of The Ottawa Hospital, with the objective of obtaining information regarding adverse events. In particular, AEMS was used at the Intensive Care Unit (ICU), at the Neurosurgery Unit, and at the Emergency Unit.

After investigating triggers and collecting data on events occurred within the inpatient population of these units, the OHRI principal investigator of the adverse event project observed that stakeholders and hospital personnel did not have reports capable of showing the results of the investigation carried in a simple way. Furthermore, there were neither templates available nor guidelines on how, when, or to whom this information should be reported.

In order to provide some feedback to the analysed units, simple reports were manually constructed by the OHRI principal investigator. Information was extracted from the database tables and inserted into Microsoft Excel spreadsheets to have the ability to generate graphical charts and to obtain aggregate data such as totals and percentages. These reports specifically contained: the total number of AEs, preventable AEs and potential AEs, as well as totals by error type and severity of harm. They were complemented with detailed descriptions for each incident and some charts to provide a visual idea of the outcomes.

The end result of this process was the creation of basic reports, which contained some crucial facts but lacked the ability to address the global objective. Much manual work also went in the creation of such reports, with risks of introducing errors or bias along the way.

Also, an analysis was made to understand the portion of the database that had been used for the reports. Its results showed that most of the collected data during the pilot implementations had not been used towards any report. This fact diminished the value of the data collection exercise.

3.2.2 Development of the Methodology Prototype

The situation described in the previous section set us to work on the development of a methodology capable of providing health care personnel with a tool to discover, select, and implement metrics which could report on the collected data (e.g., adverse events in the AEMS system).

A literature review provided a first understanding on the topic and led to valuable insight on the many challenges faced by health care researchers when trying to quantify AEs and establish metrics capable of reporting on them. One outstanding issue was the provision of direction to the measurement exercise. Many methodologies suggest following guidelines or using benchmarking tools to direct the selection (or discovery) of metrics instead of analysing organization goals and let those goals lead the measurement exercise. As a result, report-consumers end up with metrics that do not provide the information they need in order to take appropriate decisions. This problem guided our investigation towards goal-driven methodologies. However, to our knowledge, the health care sector has not used these methodologies extensively. This sector is in need of a goal-driven methodology that could be used to generate metrics which responded to stakeholder and organization goals.

As seen in Chapter 2, the software industry has used some approaches that integrate goal-driven development of metrics, in a direct manner. It was hence decided to use and tailor a software-related methodology in the context of health care metrics development.

After analysing several methodologies and comparing their steps, inputs, outputs, and potential applicability, as shown in Table 1, GQM (discussed in Section 2.1.1) was selected as the foundation for the creation of a health care goal-driven methodology. This selection was based on the following criteria and observations:

- GQM is simple and straightforward enough to be understood and used by health care researchers, doctors, nurses, and other stakeholders.
- GQM's goals/questions/metrics schema provides the guidance and goal-driven structure required in our context.
- The fact that, within the software development sector, GQM is already considered a "de facto standard for the definition of measurement frameworks" [Berender et al., 2006 p316] increased our confidence level in using it as our base approach.

Accordingly, the steps related to GQM's core are used as a base for putting together a prototype methodology, later named *Health Care Goal Question Metric* (HC-GQM), to make reference to the use and adaptation of GQM to the health care context.

Literature also provided other important factors to take into consideration, such as: constructing a team capable of pushing forward the metrics development project [McGarry et al., 2002; Veillard et al., 2005]; involving stakeholders beyond the collection of goals, questions, and metrics; using stakeholders' input to build and later validate reports; and using data warehouses to make the reporting process easier and more efficient [Kimball and Ross, 2002].

Consequently, the prototype version of HC-GQM includes the following steps:

Step1: Select team members and stakeholders to participate in the metrics development project

This step is composed of two tasks: i) the selection of a team that will guide and champion the development of the metrics, and ii) the selection of the stakeholders that will provide the information required to identify goals, questions, and ultimately metrics.

Step2: State business goals

The team should help the stakeholders focus on goals they want to accomplish. This is not an easy task and therefore it should be introduced in a brainstorming session, where all the suggested ideas are valid and taken into consideration before being cleaned up and prioritized. To motivate stakeholders, currently available reports and "example goals" can be presented.

Step 3: Refine business goals into measurement goals

This step should streamline and transform the rather generic business goals collected in the previous step into measurable ones. The objective is to obtain a set of goals that represent the main concerns of the stakeholders while following a format that makes clear what stake-

holders should attain. For example, for a business goal such as "To reduce the number of adverse events that happen at the hospital", the measurement goal could be "To reduce by 20% the number of adverse events that take place at the Emergency Unit before December 2010". This second goal includes more information and is more precise, and therefore it becomes easier to assess whether it has been achieved or not.

Step 4: Pose questions to describe the goals

Once a set of measurement goals have been developed stakeholders should ask questions to describe the goals in depth. In this way, one can capture a common understanding of what the goal means for each stakeholder. Questions also serve another purpose, which is to make it easier to determine whether a goal has been achieved. The following example shows the outcome of this step. For the goal: "To reduce by 20% the number of adverse events that take place at the Emergency Unit before December 2010", the following questions could be developed:

a. What is the current number of AEs in this unit?
b. How many AEs happened in December 2009 at this unit?
c. In which part of this unit do AEs tend to happen the most?

Step 5: Develop the metrics

The answer to the questions comes in the form of measures. Therefore there is a need to develop the metrics that will provide the measures, which in turn can answer the questions. This analysis leads the process into a quantitative level. It should be mentioned that one metric can help answer many questions.

Step 6: Collect the measures

Once the metrics have been defined they should be collected. This process might include the development of data marts or dimensional models in order to potentially gather information from different sources.

Step 7: Define the reports and their intended readers

This step help develop a set of reports containing the metrics previously identified. Reports' intended readers (i.e., stakeholders) should be taken into consideration so that the reports present the information in the correct format, to the appropriate user, at the right time.

3.3. Implementation of the Prototype Methodology

3.3.1 Selection of the Neurosurgery Unit

The series of pilot implementations for the prospective clinical surveillance approach at TOH included the Neurosurgery Unit. It was observed during the previous implementations at the Emergency and Intensive Care Units that having a system like AEMS that collects data is not enough. It is necessary to include the creation of reports capable of explaining what is happening with a unit in terms of AE occurrences. This was already acknowledged by Blais et al. [2009] when they developed AEMS, but reporting was out of scope of their project. HC-GQM can be investigated to provide a solution for the reporting issue.

It was hence decided by the adverse event project's principal investigator at OHRI to test HC-GQM prototype at the Neurosurgery Unit using AEMS as its main data source. The data collection process started in December 2009 and ended in February 2010. During this period, the methodology was implemented and reports were generated.

3.3.2 Following the Methodology Steps

Step1: Select team members and stakeholders to participate in the metrics development project

Meetings were carried with the "prospective AE surveillance initiative" leadership of TOH with the purpose of identifying personnel that could be part of the team as well as the stakeholders that should be involved. It was suggested to select nurses, physicians, as well as the head of the Neurosurgery Unit as the stakeholders. The objective was to obtain different insights when generating goals and their defining questions.

The team was composed of a project leader, who was the principal investigator at OHRI, of a project facilitator, who was a graduate student (the author of this thesis), and of a researcher (observer) nurse.

Since this was the first trial of the methodology, it was decided by the project leader not to directly involve nurses and doctors from the studied unit. In this manner, we would avoid potential future non-acceptance of the methodology. The head of the department was also not formally included but this role was taken into consideration as a report consumer.

In this way, the project leader and the observer nurse also acted as stakeholders and defined goals, questions, and metrics. These team members were however in touch on a frequent basis with the real stakeholders at the Neurosurgery Unit (so they were aware of their needs).

Step2: State business goals

The business goals were obtained from the project plan for the implementation of an organization-wide Adverse Event Management System [LeBrun, 2009].

- "To develop, test, and implement a learning model that will facilitate measurable improvements in patient care and safety, using prospective adverse event surveillance..."
- "Improve monitoring of adverse events in hospitals"
- "To move The Ottawa Hospital from AE surveillance and reporting to action as a means to learn from, prevent, and/or reduce AE's in the organization".

Step 3: Refine business goals into measurement goals

The previously stated business goals were transformed into the following measurement goals:

1. To analyze the process of AE detection in order to understand AE occurrences within the Neurosurgery Unit at TOH
2. To compare the AEs detected with the current tool with the AEs detected with the methodology currently in place at TOH.
3. To reduce AEs in the Neurosurgery Unit.

It should be mentioned that the previous measurement goals do not represent all aspects of the business goals, since this was a pilot implementation of the prototype methodology.

Measurement goals were not analysed by different stakeholders, therefore only one point of view was gathered. In addition, their construction did not really include a measurable quality attribute. However, the objective of making the goals more understandable and structured was met.

Step 4: Pose questions to describe the goals

Goal 1: To analyze the process of AE detection in order to understand AE occurrences within the Neurosurgery Unit at TOH.

Questions:

1. How many events are captured by the process?
2. How many patients are captured by the process?
3. How many patients present events?
4. How many patients present AEs?
5. Out of the total of events, how many are AEs?

Goal 2: To compare the AEs detected with the current tool with the AEs detected with the methodology currently in place at TOH.

Questions:

1. What types of AEs are collected by this process?
2. How many events are collected by the process?
3. How many preventable AEs are collected by the process?

Goal 3: To reduce AEs in the Neurosurgery Unit.

Questions:

1. What are the types of AEs? Most common ones?
2. What are the consequences of AEs in patients? Most common ones?
3. Where do AEs happen? Most common locations?
4. Who is responsible for the AEs?
5. How many preventable AEs are reported?
6. How many non-preventable AEs are reported?

Again, this step was not addressed to full satisfaction. Many questions that could have been generated were not well formulated. The main cause was the small number of stakeholders involved in this process (only the project manager and the observer nurse participated) and their lack of time to meet, analyze the goals, and describe them in the best possible and quantifiable way.

Step 5: Develop the metrics

Straightforward questions led to the generation of metrics that were simple to implement. Table 2, Table 3, and Table 4 present the relation of questions with their metrics per stated goal.

Goal 1: To analyze the process of AE detection in order to understand AE occurrences within the Neurosurgery Unit at TOH

Questions	Metrics
How many events are captured by the process?	- Total count of events registered by the system
How many patients are captured by the process?	- Total count of patients registered by the system
How many patients present events?	- Total count of patients with events
How many patients present AEs?	- Total count of patients with events categorized as AE
Out of the total of events, how many are AEs?	- Percentage of AEs over the total count of events, - Ratio of AEs over the total count of events - Total count of AE

Table 2 Metrics Developed for Goal #1

Goal 2: To compare the AE detected with the current tool with the AE detected with the methodology currently in place at TOH.

Questions	Metrics
What types of AEs are collected by this process?	- Total count of AEs by Error Type - Percentage of AEs by Error Type
How many events are collected by the process?	- Total count of events registered by the system
How many preventable AEs are collected by the process?	- Total count of AEs categorized as preventable - Percentage of preventable AEs over the total count of events

Table 3 Metrics Developed for Goal #2

Goal 3: To reduce AEs in the Neurosurgery Unit.

Questions	Metrics
What are the types of AEs? Most common ones?	- Total count of AEs by Error Type - Percentage of AEs by Error Type over the total count of events - Max of Total of AEs by Error Type
What are the consequences of AEs in patients? Most common ones?	- Total count of AEs by Severity, - Percentage of AEs by Severity over the total count of events - Max of Total of AEs by Severity
Where do AEs happen? Most common locations?	- Total count of AEs by Location - Percentage of AEs by Location over the total count of events - Max of Total of AEs by Location
Who is responsible for the AEs?	- Total count of AEs by responsible physician, - Percentage of AEs by responsible physician over the total count of events - Max of Total of AEs by responsible physician
How many preventable AEs are reported?	- Total count of AEs categorized as preventable - Percentage of preventable AEs over the total count of events
How many non-preventable AEs are reported?	- Total count of AEs categorized as non-preventable - Percentage of non-preventable AEs over the total count of events

Table 4 Metrics Developed for Goal #3

Step 6: Collect the measures

Once a set of metrics was defined, we started analyzing the data sources to identify the required data to formally define and compute the metrics. A dimensional model was chosen over a transactional model because the schemas provided by dimensional models facilitate the reporting tasks while being more efficient and easier to understand. The dimensional model was designed following recommendations by Kimball and Ross [2002], which allowed the transformation of the transactional database provided by AEMS into a reporting model.

The first recommendation is to select the business process to be modelled. In this case, the process is the "collection of information on AEs".

The second recommendation relates to the selection of a business process level of granularity. This allows the specification of the information that would be contained in a row of a fact table or, in other words, the level of detail of the fact table. For this step, we decided to set the level of granularity by transaction instead of by periods of time (snapshots of information). This choice permits to show in a row a specific transaction, i.e., events that have been rated and categorized. In short, the *Events* fact table then describes an individual (rated) event.

The third recommendation consists in the selection of the dimensions that describe each fact table row. Dimensions define in how many different ways information can be seen. Dimensions are especially useful when they are defined hierarchically. The following dimensions were considered for the metrics: severity, error type, patient, time, location, trigger type, and responsibility.

The fourth recommendation is the selection of the facts to be measured. Facts should agree in grain level. Different grain levels might point out the need of different fact tables. Since most of the operations needed in the metrics defined in Tables 2 to 4 are *counts*, the fact table can simply be fact less, that means that no other information but the foreign keys are stored in this table. However, the fact table of this model was not fact less. A dummy field (EventCount) was added to make calculations easier and faster.

Figure 4 shows the dimensional model obtained after applying these recommendations. The notion of event is at the center while the other tables (Trigger, Severity, etc.) represent the associated dimensions through which the events can be analyzed and counted.

With such dimensional model in place, it was possible to formally define the discovered metrics. Consequently, a metric such as *Total count of events registered by the system*, can be formally defined as *Select Count (EventCount) from AEMSReport_Events*. In the same way, a metric like *Percentage of AEs over the total count of events* can be formally defined as *Select (Count (EventCount) * 100 / (Select Count (*) from AEMSReport_Events)) as AEPercentage from AEMSReport_Events where EventTypeId = 1*. Note that EventTypeId = 1 indicates the presence of an adverse event).

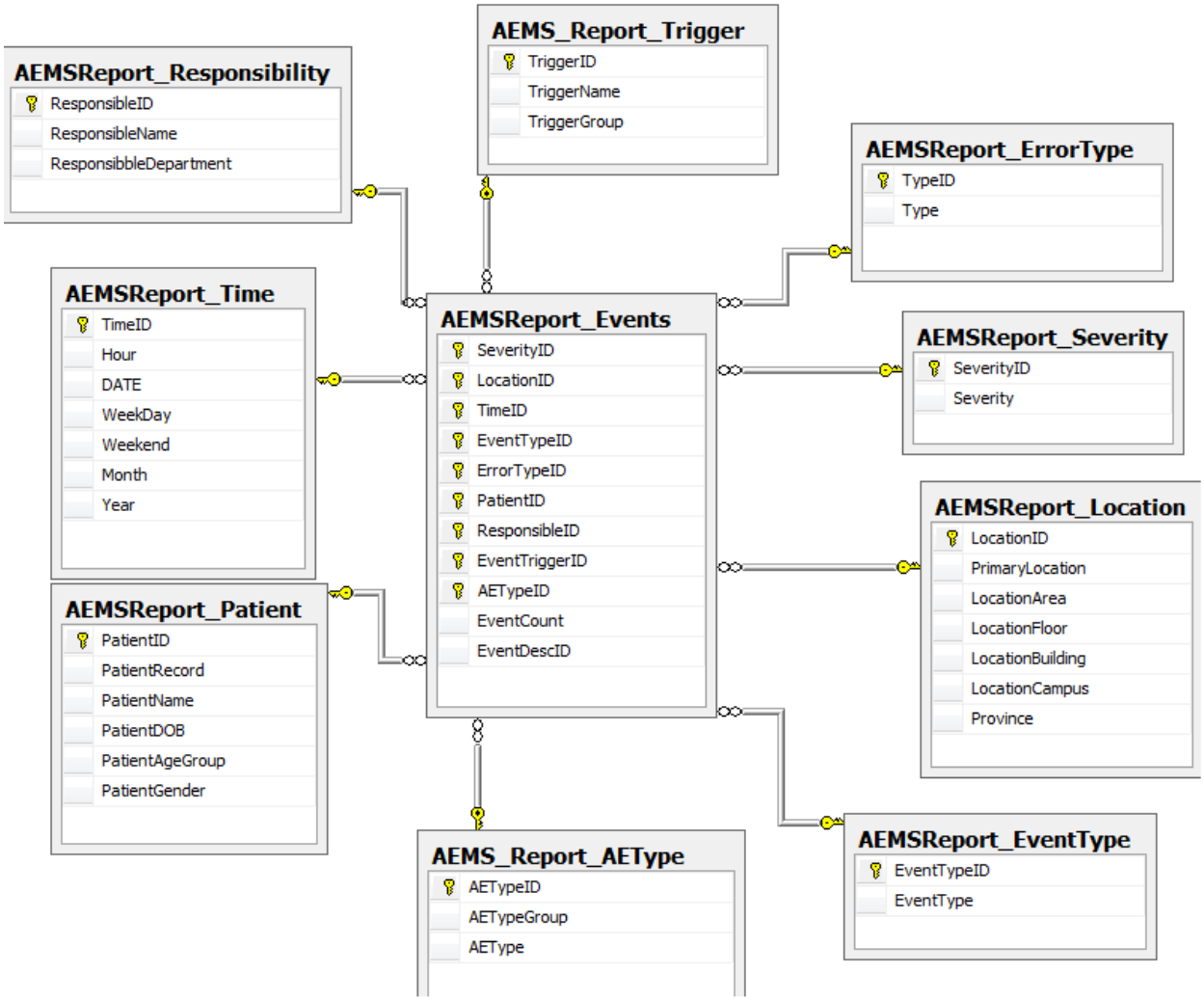


Figure 4. Dimensional Model

Step 7: Define the reports and their intended readers

This step is in charge of generating the BI reports that contain the metrics defined in step 5 using the dimensional model from step 6. In order to do so, the dimensional model in Figure 4 was implemented using the Framework Manager tool, part of IBM Cognos 8. By doing this, the construction of reports and their specificities turned into a straightforward and simple process.

The development of the metrics was handled by Query Studio, another tool of the IBM Cognos 8 suite. This application supports the building of queries that can pull information from a dimensional model based on specified criteria. Later, Report Studio (also part of IBM Cognos 8)

was used to create the reports, bringing together the developed metrics, images, graphics, and charts. Extracts of such reports (with fake data, due to confidentiality and privacy reasons) are shown in Appendix A.

The reports generated addressed concerns presented in the questions. It should be pointed out that after developing the dimensional model, the OHRI principal investigator determined that not all dimensions identified could be used in the reports. Some of them such as AEMSReport_Patient or AEMSReport_Responsibility contained information about patients or doctors which could not be disclosed due to privacy legislation. In cases like AEMSReport_Location or AEMSReport_Trigger, it was also noticed that the fields in the database had not been properly filled and therefore no meaningful information could be extracted from this data.

3.4. Difficulties Encountered

During the implementation of the HC-GQM prototype, we detected aspects that were believed to negatively influence the outcomes of the measurement exercise. Because of these issues, the methodology did not produce the expected results. In addition, even if certain reports with their associated metrics were developed, the possibility of obtaining more meaningful outputs such as trends and patterns was not achieved.

The main issues are grouped below into three main categories of methodology weaknesses, in order to enable their understanding and later to approach them with solutions:

a) Data source weaknesses:

- The information contained in the database provided by AEMS did not provide enough information to compute all required metrics and to build the all required reports.
- AEMS did not enforce the correct or complete input of information, leading to data quality issues that also limited the creation of some metrics and reports.

b) Methodology process weaknesses:

- The methodology did not enforce a mechanism to avoid a poor selection of stakeholders participating in the project: the team was formed by the principal investigator, a project facilitator and the observer nurse. The chief of the clinical unit, other physicians, and

quality assurance personnel were not included. This implementation did not benefit from the different points of view that come with larger teams. The smaller the team, the fewer perspectives will be taken into account when describing the goals in the form of questions, which in turn affects the elaboration of more complete sets of metrics.

- No action plan was conceived. Stakeholders did not have any guideline to determine when exactly a step should end and when the next one should begin. Similarly, stakeholders were not forced to present the outputs of each step in a rigorous manner. This promoted some disagreements and unnecessary delays.
- Goals, questions, and metrics were developed without taking into account the information available in the data sources; therefore the collection step of the methodology could not be completely achieved. Many metrics required the collection of data unavailable in the AEMS database.
- The creation of reports occurred at the end of the metrics development process. The presentation of reports that contained the metrics made users realize the many aspects or points of view were not properly considered when developing the metrics.
- Many questions that could have helped describe the goals in a more complete manner were not identified. This decreased the potential of generating more meaningful reports. Example of such questions are:
    - When do AEs tend to happen the most? Days or night? Week days or weekends?
    - What are the most commonly used triggers?
    - What is the average time between an event's collection and its categorization?
    - What is the percentage of patients that present AEs?
    - How long does it take to categorize an event?

c) Team members' weaknesses:

- Lack of time of participating stakeholders: Not only was the team small but its members generally lacked the time to meet in order to work together on the methodology.
- Team members had vague ideas of what information they wanted on their reports, and this got reflected on the questions asked, and later on the metrics generated.

3.5. Conclusion

This chapter provided the details of the health care context that led the research to focus on the creation of a goal-driven methodology for metrics development. A prototype methodology (HC-GQM) was generated so it could be used by health care personnel to develop metrics capable of reporting on adverse event data collected by the AEMS system.

The implementation of the seven steps of the HC-GQM prototype methodology in the Neurosurgery Unit of The Ottawa Hospital and the results obtained led to observations on three categories of weaknesses related to data sources, to the composition of teams, and to the methodology steps themselves. Some of these issues are addressed in the next section, with a refined version of HC-GQM.

Chapter 4. Refining HC-GQM

This chapter explains how, by analyzing the methodology difficulties reported in the previous chapter and by performing a deeper literature review, it is possible to enhance the HC-GQM prototype to address these difficulties.

Section 4.1 presents a deeper analysis of reported weaknesses of the methodology previously prototyped. It also discusses where HC-GQM can be amended in order to solve the problems identified. Section 4.2 proposes a refined HC-GQM methodology, which is then explained in detail in Section 4.3.

4.1. Analysis of the Reported Difficulties

By the end of the implementation of the prototype methodology, some difficulties were detected and reported in three main categories: data source, methodology, and team. This grouping reflected the respective natures of the issues.

In this section, a deeper analysis of the reported weaknesses is performed with the objective of separating those that directly influence the methodology's performance from the ones that are associated with external factors.

4.1.1 Data Source Issues

This category contains weaknesses related to the application that was used for data collection purposes. In order to process the data source provided by AEMS and make it usable for reporting services, it was necessary to standardize the data contained in the fields of the database. This process deleted duplicated information and events that had been introduced as "test cases", i.e., fake events. It also made sure that no patient identifiable data was available and finalized the classification of some events, among other tasks. This process of *standardization* (or data cleaning) is quite common when moving records from a transactional database to a data warehouse. However, some difficulties were faced for which the standardization process could not do much:

- the small number of records related to patients and events collected;

- the occurrence of incomplete fields within the records;
- the non-existence or non-collection of required information.

a) Small number of records related to patients and events collected

This issue reduces the array of possible goals to be addressed. If there is not enough data to work with, then metrics such as trends cannot be developed. In the same way, business intelligence capabilities such as the "forecasting" of certain behaviours cannot be supported. Moreover, by having just a small number of records to be analysed, the validity of certain metrics that provide statistical information is negatively affected.

b) Occurrence of incomplete fields within the records

The fact that data was either incorrectly introduced or not introduced at all in some of the fields of the database made the process of obtaining meaningful metrics harder. Possible data "associations" could not be constructed. For example, the location where the events happened was not properly introduced for most of the events. This situation weakened the answers to questions such as "How many patients suffered a *fall related event* at a certain *location* of the hospital". In this case, the association of "fall" events with a "location" could not be completed. The absence of event location information in the system constrained the output of the application of the methodology.

c) Non-existence or non-collection of required information

The fact that "audit-type" information was not collected also reduced the number of goals and questions that could be generated. For example, events change their status as they go through the process of being identified, to being categorized and then closed. By not keeping a list of dates and times of when these statuses changed, it was not possible to address goals related to the performance of the prospective surveillance method. Questions such as: "How fast are events registered in the system?" or "What is the time span between an event's capture and its processing and further classification?" could not be answered.

Furthermore, since the available information was only related to one clinical unit within one campus of the hospital, metrics that could compare the same unit across different campuses

of the hospital could not be generated. This influenced negatively the results since the unit could not compare itself with other units and know how to interpret its results.

Although all of the above three difficulties constrained the possible assortment of outputs of HC-GQM, (i.e., the number of metrics and reports), they are not strongly related to the methodology used. In other words, there is nothing HC-GQM can do when the data sources used do not provide the required information, except promoting the collection of the missing data by showing how it could be used.

4.1.2 Methodology Process Issues

The difficulties collected in this group have a strong relation with steps of the methodology that were either poorly performed or non-existent. They are:

- lack of an action plan;
- poor knowledge of the data source;
- creation of reports at the end of the exercise.

a) Lack of an action plan

This issue exposes the consequences of not having an action plan to provide to the stakeholders and team members. Action plans are essential to the management of processes and provide a road map of the activities that need to be performed, important deadlines, deliverables, etc. By not providing an organized plan, the team members did not know what activities were supposed to happen, when or who was supposed to participate. Under this situation, team members could not arrange their schedules, visualize the needs of the project or their possible results. This promoted disagreements between the team members and added unnecessary delays to the completion of the project.

Tentative solution: In order to address this difficulty, it is necessary to add some level of management to the methodology. The new solution should have a plan in place exposing the main objective of HC-GQM, the steps that stakeholders should go through, what should be accomplished by the end of each step as well as timelines, and a distribution of responsibilities among the team members.

b) Poor knowledge of data source

The second methodological issue is related to the reality that goals and questions were developed by the stakeholders without taking into account the "facts" stored in the available data source. This difficulty is highly related to both other categories: "data sources" and "team members". Since the project did not have a data provider person as part of the team, it was not possible to know the details of what was actually stored and available for use in the data warehouse. This apparently simple issue brought along many other difficulties that ended up lowering the team morale and making them feel confused and frustrated. The following examples illustrate the consequences of this problem.

Team members went through the steps of the methodology and developed a set of goals, questions, and metrics. However, when trying to collect the information from the data source, they realized that this was not possible since the requested information was either incomplete or corrupted. Goals and questions related to where the events were happening were left unfinished. Similarly, AE rates within inpatient population were also incomplete. Note that this issue actually resulted in more training of the observer nurse, to ensure that the appropriate information would actually be collected in the next usage of AEMS.

In other occasions, the team created goals and develop their related questions. When metrics had to be built, they realized that such questions could not be responded to because the system (AEMS) was not collecting that kind of information at all. In this case, questions such as "What is the average processing time of an event?" were also left unanswered.

Tentative solution: In order to solve this situation and avoid other conflicts, a data source specialist who can provide insightful information regarding the state of the data needs to be included. This individual would be valuable in communicating to the team what data is available as well as analyzing the possibility of bringing together diverse data sources in order to provide the info that stakeholders need. This does not prevent creating metrics that cannot be computed right away (as the required information may become available later, with updates to the data collection process), but this would help managing expectations.

c) Creating reports at the end of the exercise

The third difficulty contained in this category relates to the reports development phase. Their creation occurred at the end of the metrics development step. Only then, team members were

faced with the outcome of their work, i.e., a set of reports providing data, with graphics and charts that helped them understand the real state of their unit in terms of AEs. Although valuable, these reports were not completely in synch with stakeholders' needs as they did not contain the information in the exact format stakeholders wanted. Furthermore, stakeholders realized by looking at the data that many aspects were not considered during the creation of goals, questions, and metrics. As a consequence, new goals and questions addressing different perspectives came to light.

The fact of having a tangible result, something to look at, analyze, and criticize, helped them rethink some of the goals and questions previously posed.

Tentative solution: The above difficulties could have been avoided if mock-up reports had been constructed while developing questions and metrics. From a stakeholder's perspective, reports are the ultimate and most desired outcome of the methodology. They contain the data they want in the format they need. Reports show the result of their work and therefore their development should have been a progressive and iterative work. The building of report prototypes as part of the methodology can help save time and effort.

4.1.3 Team Members Issues

The difficulties gathered in this category refer to the size, diversity, and commitment of teams. They have been classified as:

- team composition;
- team diversity;
- team members with busy schedules.

a) Team composition

The team was composed of only three members (a project leader, a project facilitator, and the observer nurse). They played several roles that should have been allocated to other people. For example, they defined the project goals (upper management role), developed measurement goals (measurement team role), defined questions (measurement team with data providers), etc.

b) Team diversity

All team members that acted as stakeholders were part of a same unit (OHRI researchers). This poor diversity brought as a consequence that different points of view were not available and therefore not considered. Most of the defined goals and questions revolved around clinical issues whereas, for example, risk management and quality assurance perspectives were not included in the analysis.

c) Team members with busy schedules

The members of the team who acted as stakeholders (principal investigator and observer nurse), had busy schedules and little time for meetings. This issue has also hurt the methodology in that, often, meetings had to be postponed because the team could not get together, and the time separating meetings became long enough to forget the outputs obtained in previous steps.

To summarize, these three difficulties reduced the possibilities of finding strong, consistent, and diverse goals, which in turn compromised the process of questions generation and finally the meaningfulness of the metrics obtained.

Tentative solution: All of the above weaknesses directly affect the methodology, its performance, and ultimately its outcome. It is not hard to foresee the results of choosing the wrong team: negative impacts in the approach and its results. Therefore, there is a need to modify the methodology in order to select team members who belong to different areas and are willing to contribute. It is also important to gather enough members in the team so that the roles are well distributed among the members. Furthermore, the project team should include executive managers of the organization. In this way, they can make the measurement exercise a priority, thus forcing the rest of stakeholders to create time for meetings and other related activities.

Table 5 presents a summary of the issues and their possible solutions.

Category	Issue	Solution
Data Source	Small number of records	No solution provided
	Incomplete fields within the records	No solution provided
	Non-collection of needed information	No solution provided
Methodology	Lack of an action plan	Creation of an action plan as part of the methodology
	Poor knowledge of data source	Include data source specialist in the team
	Creating reports at the end of the exercise	Create report prototypes in an iterative way
Team Members	Team composition	Modify the team creation step to select members in such a way that they do not hold several roles in the project
	Team diversity	Modify the team creation step to select member from different areas
	Team members with busy schedules	Include executive managers as part of the team

Table 5 Summary of Issues

4.2. Resulting Goal-Driven Methodology: HC-GQM

The analysis of the above difficulties was a necessary step to understand their nature and characteristics. Furthermore, this analysis promoted the development of possible solutions for each issue, which included the generation of new steps and variations to existing ones. Consequently, this process brought a reorganization of HC-GQM to accommodate the latest changes (Figure 5).

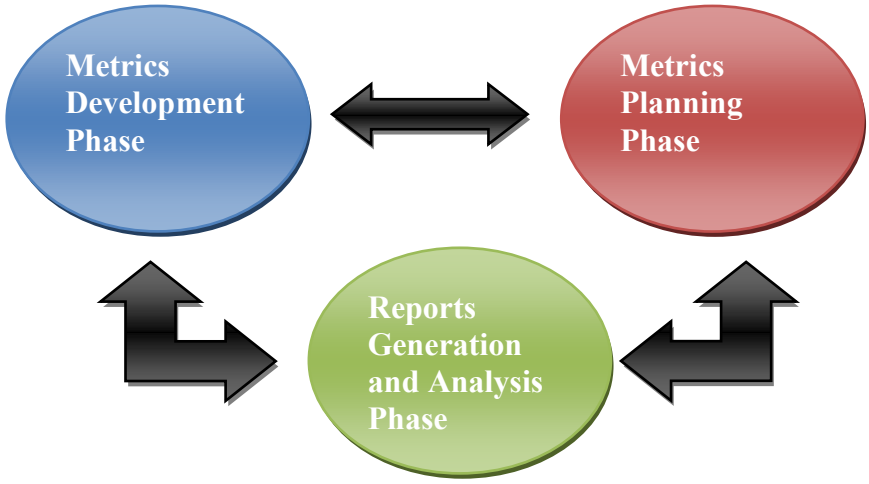


Figure 5. HC-GQM Phases

The refined HC-GQM is composed of three interdependent phases that communicate and retro-feed each other continuously: Metrics Development, Metrics Planning, and Reports Generation. It should be emphasized that one phase does not need to be completed for the others to start. Actually, the Metrics Planning Phase expands along the project's life and interacts with both the Metrics Development Phase and the Reports Generation and Analysis Phase. By grouping the methodology's tasks into those phases, it becomes possible to isolate their purpose and run them simultaneously. At the same time, their independence allows the project to adapt and transform according to the characteristics of the environment while incrementally progressing. It is important to state that the methodology should be used iteratively. This means that all the metrics and reports do not have to be developed in a first attempt. In this way, stakeholders and team members can get familiar with the methodology, its steps, and its outcomes at the same time that they evaluate the results of the first iteration. This model enables them to improve the reports and the metrics created.

Figure 6 illustrates, in a more precise manner, the interactions among the three phases of HC-GQM.

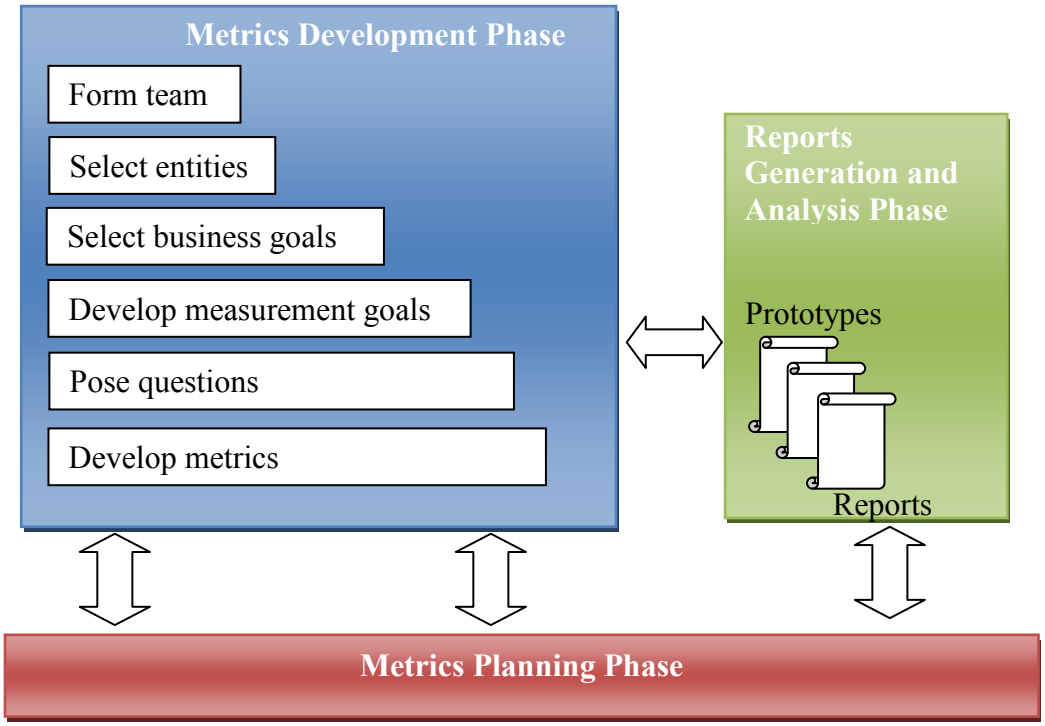


Figure 6. Interaction Among HC-GQM Phases

The Metrics Development Phase includes many of the steps of the initial methodology: forming a team, generating goals and questions, and creating metrics. Although these new steps might have the same names as the old ones, they have undergone modifications to reflect the lessons learned during the first implementation. The activities included in this phase should be carried out following a top-down approach. However, it should be noted that this linear approach can be modified in cases where the outcome of a certain step leads the team to go back to previous steps. For example, step 5: "Pose questions" can lead to the discovery of new goals. In such cases, the team addresses steps 3 and 4 again. Also, the Metrics Planning Phase should perform the necessary changes to include newly discovered goals in the plan. This will be further addressed in Section 4.3.1.

The Metrics Planning Phase is present from the beginning to the very end of the project. This phase handles the project management tasks that are needed to carry the project forward. It delimitates the starting and finishing points of HC-GQM's activities and it provides guidelines on the project deliverables to be created. While most of its workload resides at the beginning of the project, the coordination effort that this phase demands alongside the project's life span is essential to the success of the metrics development exercise. This will be discussed in detail in Section 4.3.2.

The Reports Generation and Analysis Phase has two different purposes: the construction of reports and the analysis of the information collected. This phase is initiated once the measurement goals are defined. While the Metrics Development activities are performed, the report prototyping should also be carried out. This then facilitates and enhances the stakeholders' visualization of their work while validating each activity performed during the Metrics Development Phase. The analysis of the results occurs after the reports are developed and the measures collected. By studying the results, it is possible to provide answers to the posed questions and in this way address the stated goals. More information on this topic will be given in Section 4.3.3.

4.3. HC-GQM Phases and Activities

4.3.1 Metrics Development Phase

This phase is the core of HC-GQM since it provides a series of steps that support the discovery, development, and implementation of metrics. It contains the following steps:

- Form team
- Select entities
- Select business goals
- Develop measurement goals
- Pose questions
- Develop metrics

Step 1 – Form team

HC-GQM starts by forming a team of individuals who have a particular interest in the measurement results. In order to avoid the difficulties previously explained, the team should contain the roles explained below. Although one person can assume more than one role, it is recommended to have different people filling each position. The diversity of criteria that different backgrounds offer tends to enrich the overall experience. It should be noted that stakeholders are also part of the team as they are key to the development of goals and questions. Stakeholders can be found in the roles of organization managers and quality assurance personnel. The team should therefore be formed by the following roles:

- Measurement Manager: The person holding this position is the team leader and is in charge of putting together the action plan, acting as a mediator between the interacting parties, scheduling meetings, and keeping all stakeholders motivated. The Measurement Manager knows each of the steps to be accomplished and their requirements. He or she also manages timelines and assigns tasks to the team members.
- Organization Manager: It is necessary to engage several levels of management across the organization in order to obtain appropriate guidance and direction. By involving executive management, the measuring exercise gets the appropriate support, thus minimizing delays due to unavailability of its members. The project is also provided with high-level goals that the organization wants to address. However, executive management it is not sufficient. It is also important to engage project managers since they can define more specific goals related to their projects and areas of expertise.
- Quality Improvement Personnel: One of the many purposes of implementing a measurement program within organizations is to improve quality. Therefore, it makes sense to

have specialists who can guide the team towards developing the goals and asking the questions that will better address quality improvement.

- Data Provider: The function of this role is to provide the team with knowledge regarding the availability of and access to data. The data provider is able to determine what can be realistically collected, therefore providing the team with objectivity while managing expectations.
- Statistician: The metrics need to lead to reliable and valid results. Simple mathematical equations do not always reflect the empirical world in the way they should. There is also the danger of creating relationships among entities and their attributes that do not really exist. Therefore, it is imperative to have a subject matter specialist who can take this step further and provide valid metrics for the measurement exercise.

Step 2 – Select entities

Entities represent "what" will be measured. The purpose of this selection is to streamline the measurement process by reducing a potentially wide array of possibilities to very specific ones. In other words, by selecting entities, stakeholders focus on determined processes, areas, dimensions, and systems that they want to study and from which to obtain information.

During the implementation of the prototype HC-GQM, it was noticed that if stakeholders were provided with a certain *entity* (e.g., a process) to focus their attention, it became easier to develop the goals, questions, and metrics needed to address this entity. By not trying to embrace all measurement needs at once, stakeholders can actually go through the methodology steps without getting confused or overwhelmed.

Step 3 – Select business goals

Business goals must exist to establish measurement goals. Without them, the discovery of metrics has no focus. This task requires that the organization managers and quality improvement personnel meet to discuss the needs and expectations of the measurement exercise. The previous experience showed that this task cannot be easily accomplished without some help. Stakeholders need a guide to start the development of their goals. We learned that the analysis of existing reports, although useful, is not the tool that stakeholders require.

After evaluating ways of providing help to stakeholders, we suggest that a literature review be carried alongside the methodology steps. Several researches [Greenberg et al., 2005] [Veillard et al., 2005] have used such tool in order to gain knowledge about what has been developed and validated by others.

Therefore, we recommend that the measurement manager performs a literature review in order to find goals that address the set of entities to be measured. Such review permits to tailor goals that are "already used" by similar organizations or imposed by governments.

Step 4 – Develop measurement goals

Business and measurement goals are not always mutually exclusive. Measurement goals describe how to track the progress of business goals. This description, suggested by Basili et al. [1994], is accomplished by including what needs to be measured, by whom, under what context, and with respect to what.

The team should perform a brainstorming session in order to facilitate this process. The template in Table 6 can be used for this purpose [van Solingen et al., 1999]. The left column usually remains constant on the form while the right column changes for each goal defined:

Facets of Information [van Solingen et al., 1999]	Example
Analyze *(the object under measurement)*	the Adverse Event (AE) detection mechanism
For the purpose of *(understanding, controlling, improving...)*	reducing the number of preventable adverse events by 20% in 2010
With respect to *(quality attribute of the measured object)*	location where AEs happen
From the point of View of *(stakeholder involved)*	the Unit Manager
In the context of *(environment)*	the Neurosurgery Unit during working hours on week-days

Table 6 Template Table for Goals

Step 5 – Pose questions

This step has not changed much since the version proposed in Section 3.2. As previously described, questions need to be posed to address the *perception of quality* and *the context or environment* of the goal.

At this point, all stakeholders should converge on a common understanding and interpretation of the goal in the given environment. To accomplish this, the measurement manager can either perform individual interviews or group meetings. The mechanism is chosen depending on the availability of the team members. However, it is recommended to perform group meetings since this encourages discussions and broadens the stakeholders' perspectives. In some cases however, the resulting questions may be too generic, and this hinders the appropriate identification of metrics. In such situations, questions should be divided into sub-questions. In any case, by answering the questions, one should be able to conclude whether a goal is reached or not.

According to the previous experience while implementing the methodology, the process of generating questions might lead to the discovery of new goals or to the refinement of existing ones. New goals mean that new questions will have to be posed. It is obvious that this step promotes the iterative nature to the framework. This should be taken into consideration in the planning phase of HC-GQM.

Some sample questions for the following goal are described below.

Goal: *To analyze the AE detection process for the purpose of reducing the number of preventable adverse events by 20% with respect to the location where the AEs occur from the viewpoint of the project manager in the context of the neurosurgery unit during working hours on week-days by December 2010.*

Example Questions:

- *What is the current number of Preventable AEs?*
- *What is the rate of Preventable AEs?*
- *What is the percentage of Preventable AEs with respect to the total number of AEs?*
- *Where are most of the AEs taking place?*

Step 6 – Develop metrics

Section 3.2 argued that metrics provide the quantitative information required to answer the posed questions in a satisfactory way. In this context, many significant metrics can be found. However, this does not mean that all of them should be used. Only the key metrics should be developed and then collected.

The process of developing metrics can be complex and difficult. At this stage, a literature review can help finding standard metrics frequently used to measure the selected entities (process, dimensions, etc.). However, this should be done once the questions have been developed. If a literature review related to the metrics occurs before this point, it could potentially bias the definition of goals to match already existing and valid metrics.

To create metrics it is necessary to have statisticians, data providers, and quality improvement personnel involved, since they are the experts on the topic. Furthermore, it is recommended to follow certain criteria when choosing metrics from the literature or creating new ones. The following questions can serve as a guide to develop the metrics needed:

- Is this metric valid and acceptable by the potential users? Is it valid in different contexts?
- Is it important and relevant to the current context? Does it cover aspects that matter to the users?
- What is the potential for use? Is it possible to start an action if the metric reveals an action?
- How hard is it to collect?

Many of these questions have already been used by Veillard et al. [2005] and others to reduce the number of indicators found in an initial list. We believe that if these conditions are met before selecting the metrics, the results will be closer to what is really needed.

However, as suggested by the literature [Berender et al., 2006; Veillard et al., 2005; Greenberg et al., 2005], no matter how good the initial list of indicators looks, the selection process should be performed and the list quite possibly narrowed down. This is best done by experts who develop a ranking system in order to prioritize one metric over another, ultimately choosing those that are best suited to the criteria previously presented.

After the selection of metrics, it is important to formally define them. For this purpose, a table-like template can be used. We recommend seeking the help of bio-statisticians or other ex-

perts in the field who participate in the measurement initiative. Table 7 serves as one possible example of a way in which to group metrics successfully. We also advise that improvements can be made on this table once it has been tested as a trial to show any shortcoming that may exist.

Metric	Description	Definition	Data Source
Total number of Preventable AEs by Location	Will count the number of preventable AEs in each location	COUNT(Prev_AE)	AEMS DB
Percentage of Preventable AEs by Location	Percentage of preventable AEs by location with respect to the total number of AEs	%Prev_AE = (*Total number of Preventable AEs by Location * 100)/ Total number of AEs*	AEMS DB

Table 7 Template Table for Metrics

4.3.2 Metrics Planning Phase

The Metrics Planning Phase outlines the management of the project. This phase was introduced in order to effectively address many of the difficulties presented in Section 4.1. Furthermore, it allows the formal definition of the project scope, which is not to be confused with the goals development steps that address quality concerns. This scope addresses operational issues.

In this phase, the first step is then to set the scope. This means delimitating the project and setting out an objective, before engaging in the planning phase. This activity will be mainly run by the measurement manager together with organization managers.

Once the scope is identified, a team must be selected. To put together an efficient team, the guidelines stated in step 1 of the previous section should be followed.

Only then can a plan be created. This plan must contain a calendar with activities and scheduled timelines and it must specify the team members who will be responsible for the planned deliverables.

Creating the scope, selecting the team members, as well as planning and identifying a calendar will be the most time-consuming exercises in the Metrics Planning Phase. However, these are precautionary steps that are worth following as they will provide the project with a shape and a purpose, along with an expected timeline for deliverables and distributed responsibilities.

These initial steps are what spearhead the planning phase at a later date. As the project develops however, the initial plan must be updated in order to remain synchronized with the ac-

tual execution of the project and the reality of the environment in which it takes place. Meetings can be changed, responsibilities may be switched, and deliverables potentially postponed, but close adherence to the calendar is advised in order for the project to progress within an acceptable timeframe. Also, at the end of each activity, the outcomes should be registered in a document. In this way, the results are formalized and the activities closed, and the project can move on.

4.3.3 Reports Generation and Analysis Phase

This phase was created in order to address some of the weaknesses reported in **Error! Reference source not found.**. It outlines two important processes: the creation of reports and the analysis of the project's results. The flows of information for each process follow opposite directions. In other words, while developing the metrics, the flow follows a top-down approach. This means that the reports prototyping starts when the goals are set and ends after the development of metrics. On the other hand, the analysis process follows a bottom-up approach in that once reports are created, the examination of the measures they contain provides an answer to the developed questions. This in turn allows the tracking and analysis of the state of the goals.

The creation of report prototypes is used to obtain feedback from stakeholders while keeping them interested and involved in the project. This is only possible if, while creating the prototypes, we target the stakeholders who developed the goals and questions that the report is addressing. This promotes the flow of activities and keeps the exercise evolving and moving forward.

The report prototyping should start simultaneously with the Metrics Development Phase. If reports evolve together with goals, questions, and metrics, the team is then capable to expose in a visual manner the intent of each of these steps. This process facilitates the quick gathering of feedback and the application of corrective actions if needed. Depending on the complexity of the goal and the questions involved, the number of reports per goal may vary.

4.4. Conclusion

This chapter performed an analysis of the difficulties encountered while implementing the prototype HC-GQM at the Neurosurgery Unit of The Ottawa Hospital. This analysis focused on discovering the causes of the issues and proposed solutions to address them.

As a result, a refined HC-GQM methodology was elaborated. Its improvements include the creation of three phases that interact with each other while incrementally and iteratively reaching their final objective, which is the generation of metrics and reports capable of addressing well-defined goals.

The next chapter targets the validation of this refined methodology by using it on a second health care case study, namely a Patient Safety Learning System.

Chapter 5. Implementing and Validating HC-GQM in the Context of PSLS

Chapter 5 focuses on the implementation and validation of the revisited methodology. This process takes place at TOH yet it uses a different application to collect the required information. Section 5.1 gives an overview of the selected application: the Patient Safety Learning System (PSLS). Section 5.2 elaborates on the implementation of HC-GQM's three phases for PSLS metrics and reports. Section 5.3 then presents validation results with an emphasis on the answers to a survey completed by the team members and stakeholders involved.

5.1. Patient Safety Learning System (PSLS)

5.1.1 The Need for Generic Incident Reporting

In earlier chapters, the Adverse Event Management System (AEMS) was introduced. As described in Section 3.1, this system implemented a prospective adverse event surveillance method. AEMS was used by OHRI to collect data regarding AEs. Pilot projects took place across several services of TOH such as: Obstetrics, Internal Medicine, Emergency, Cardiac Surgery, Orthopedics, ICU, and Neurosurgery. Although this application served its purpose as a pilot tool, it was not generic and scalable enough to be used in all departments and services of the hospital. Moreover, the fact that data could only be gathered using prospective surveillance also diminished the system's value as an official, institution-wide tool. For example, Accreditation Canada (formerly known as the Canadian Council on Health Services) imposes requirements regarding the implementation of a self-reporting system for accreditation purposes [TOH, 2010]. The self-reporting and the general reporting of incidents (including adverse events), by hospital personnel and even by patients and visitors, is complementary to prospective surveillance, and hospitals in Ontario are legally required to support such systems.

In order to comply with Accreditation Canada regulations, TOH had an existing institution-wide implementation of a self-reporting system. However, the hospital realized this was a rudimentary, paper-based approach that promoted apathy as well as finger-pointing among hos-

pital personnel, who were conscious of receiving the blame for the AEs that occurred. This system was also inadequate, as self-reporting mechanisms do not detect 90% of events classified as AEs [Gurwitz et al., 2000]. In addition, since this system was paper-based, it was very difficult to get meaningful reports with measures demonstrating how the hospital was performing.

TOH also needed to gather information regarding incidents in general. As described in Section 1.1, incidents are a far more generic type of event than AEs. Fires, complaints, security issues, law suits, thefts, etc. are all considered incidents. They might cause or have the potential to cause damage to patients, staff, visitors, equipment, or any property for which the institution is responsible, hence increasing the level of risk to the hospital.

In order to monitor this risk, TOH needed to collect and report data not only on events that directly or potentially cause harm to patients but also on incidents. Indeed, some of the TOH departments had already developed and adopted such systems. This was the case for Radiation Oncology with their "Incident Learning System" (ILS), Obstetrics and Gynaecology with the "Quality Indicator Notification" (QIN), and Surgery with the "National Surgical Quality Improvement Program" (NSQIP).

The approaches used in different units varied greatly in terms of methodology, reporting-techniques, and objectives. There was no electronic system capable of monitoring AEs and incidents in the institution as a whole. To solve this difficulty, TOH considered the pros and cons of various solutions. They questioned whether a strict and standardized institution-wide tool made more sense, or whether each department needed to have a system tailored to their individual needs. It was determined by TOH executives that a combination of both was best suited for their organization.

5.1.2 New Datix-Based Patient Safety Learning System

The Ottawa Hospital selected the *Datix* system [Datix, 2010] to provide a solution to the incident reporting problem. Datix, having successfully implemented this type of system in hospitals in British Colombia, Alberta, UK, and elsewhere in the world, is a reputable company that offered an application that could be tailored to the needs of TOH, especially in terms of prospective surveillance.

The system provided by Datix is composed of several modules that can handle the management of AEs and incidents. This compound solution was already named "Patient Safety

Learning System" (PSLS) in both British Columbia and Alberta. It was believed that this name was generic enough to avoid confusion between the terms AE and incidents while re-enforcing the goals of the initiative: to provide safety and to learn from errors.

PSLS is a web-based application that combines several methods for discovering and reporting on AEs and incidents. It not only provides the voluntary (or self-) reporting method requested by Accreditation Canada, and the prospective surveillance method tested by OHRI, but also a more innovative and automatic technique based on electronic triggers. Such triggers represent a promising tool to report on inconsistencies within the data collected, without the help of human involvement.

PSLS is a system that can be tailored to the needs of the stakeholders. It offers many options and variables that can be adapted to the specific criteria sought by particular departments within a larger institution. This flexibility was precisely what TOH wanted to reconcile the differences between departments with the need for a common interface used by all.

In terms of risk management, PSLS offers reporting capabilities on AEs and incidents to allow the institution to not only understand the risk of the individual departments but also to assess potential future risks to the institution and its assets. Further, PSLS also provides users with a plethora of potential reports based on criteria selected by the user. The Datix website explains the benefits of "trend analysis" and customized "detailed reports on each incident for staff on wards, statistical breakdowns for the risk management committee, or graphical analyses of trends for the Board" [Datix, 2010].

For all of the benefits that PSLS brings to TOH, there are also drawbacks for which HC-GQM can offer solutions. Besides the many built-in reports that PSLS provides, this application presents users with a wide range of capabilities to build their own reports. This facilitates the customization of the tool to the users' exact needs, but this also begs for questions such as: "What precisely should we report on?", "How is this report going to help me improve my work?", "I ended up with 3 reports, which kind of tell me the same thing... How I can combine them?", etc. Such questions tend to overwhelm users and bring confusion.

Another issue is related to patient privacy and confidentiality. Under the Quality of Care Information Protection Act (QCIPA), disclosing the ratings of individual cases is prohibited [Government of Ontario, 2004]. It is however acceptable to disclose reports with aggregate data. PSLS is set up in such a way that allows for a "drill down" mechanism that exposes information

relating directly to individual cases, patient by patient. As per QCIPA regulations, this is not allowed.

By following HC-GQM, it is possible to address the issues mentioned above. The methodology users develop goals that will guide them in generating appropriate metrics and meaningful reports, with a proper level of aggregation.

5.2. Application of HC-GQM

This section details the implementation of HC-GQM, which took place at TOH between June 2010 and November 2010. It involves two different groups of stakeholders: the Risk Management group, representing corporate-level stakeholders, and the Internal Medicine Clinical Managers, offering the perspective of care-related personnel. The stakeholders were chosen to validate the methodology's usefulness to address these different perspectives and subsequent goals.

The importance of this section is paramount to this thesis as it is part of the validation process of the methodology, the other part being the qualitative survey discussed in Section 5.3. Furthermore, as the implementation was set in a real life context, the methodology was not only applied in a generic way but rather improved upon with suggestions and lessons learned throughout our meetings.

5.2.1 Metrics Planning Phase

This phase started by engaging personnel from TOH who are involved in the implementation of PSLS at the institution. In realizing the importance of identifying key reports and their corresponding metrics, PSLS managers agreed to adopt the HC-GQM methodology as a way of gaining such reports out of the newly installed Datix system.

Project Initiation

The first step to take place was the project initiation and it included setting the scope of the implementation project, organizing a team, creating an action plan, and obtaining approval from executive management. It should be pointed out that within the project initiation, step one of the Metrics Development Phase (Form a team) was accomplished.

a) Setting the project scope

Based on preliminary meetings with the hospital's director of Clinical Quality and Performance Management, the PSLS project manager, as well as the PSLS business analyst, the scope of the measurement project was identified. It was decided to run two different test cases in parallel in two different departments of the hospital. Test Case 1 would gather requirements from the *Risk Management Group* department, generally interested in the collection and analysis of incidents and AEs from the corporate perspective. This meant getting the bigger picture and being able to compare campuses, departments, services, etc. Test case 2 would take place at the level of *Clinical Managers*, specifically those attending the units of *Internal Medicine*. This second group of stakeholders is mainly interested in AEs at the unit level.

b) Organizing a team

Once the scope of the project was defined, a team had to be put into place. Not all members of the previous panel were engaged in the measurement exercise, mainly because of their general unavailability and busy schedules. Yet, we were able to get individuals to fill the roles identified in Step 1 (form team) of Section 4.3.1. The resulting team ended up being composed of the following stakeholders:

- 1 measurement manager (the author of this thesis),
- 1 project manager (PSLS project manager),
- 1 business analyst (PSLS business analyst),
- 1 data analyst (OHRI data analyst),
- 1 biostatistician (OHRI biostatistician),
- 1 risk management director (stakeholder),
- 2 risk managers (stakeholders),
- 2 internal medicine clinical managers (stakeholders).

It should be mentioned that there was no overlap with the members of the team who participated in the pilot study, except for the author of this thesis.

c) Creating an action plan

The Metrics Planning Phase also calls for the development of an action plan that can keep stakeholders engaged as well as informed and updated with the project's latest advances and changes.

The action plan was started in June 2010 and evolved during the 5 months that the exercise lasted. As part of this evolution, the plan incorporated different steps, some of which being modified along the way (see Appendix B for a snapshot of the action plan). Each change was properly documented and described in the appropriate section. This means that changes related to the Metrics Planning Phase are described in the current section while modifications performed to the Metrics Development Phase are explained in Section 5.2.2

d) Obtaining approval

The Metrics Planning Phase had to include an additional step requested by the director of Clinical Quality and Performance Management. This extra step was necessary because the director wanted to analyze and approve the tentative action plan. In this way, he could guarantee the engagement of all stakeholders as well as make sure that no hospital policy was being violated.

5.2.2 Metrics Development Phase

The implementation of this phase started in July 2010 after having a team organized (Step 1 of HC-GQM and detailed in the previous section), an initial plan of action defined, and the scope of the project outlined.

Step 2 – Select entities

According to HC-GQM, the next step to take place is the selection of entities. These entities represent specific processes, products, or resources from which data is obtained for the measurement exercise.

The measurement manager, PSLS project manager and PSLS business analyst got together to select the relevant entities. After analyzing the project, it became clear that these had already been tentatively formed within the scope definition in the Metrics Planning Phase. PSLS and the information that it collects were going to be at the center of the measurement study.

In the case of the Risk Management Group, the interest was mainly in targeting incidents because they increase various risks to the hospital. Furthermore, the reporting of incidents is required by the hospital. The AE process, however, was still relevant to this group. On the other hand, Clinical Managers, whose main focus is on patient care and safety, were more interested in

researching AEs. PSLS was hence selected as the main entity by both groups. To summarize, the project focuses on the occurrence of incidents and AEs within the PSLS context.

Steps 3&4 – Select business goals and develop measurement goals

For our two main stakeholder groups (Risk Managers and Clinical Managers), separate meetings occurred in order to elaborate the business goals that they wanted to address. This process was necessary in order to understand the various quality issues and perspectives that needed to be taken into consideration. The objective was to retain a set of appropriate metrics and meaningful reports capable of providing the information they needed in a format they would understand.

Two separate and independent meetings were planned for each group of stakeholders. The first meeting was held to help stakeholders think about what exactly they needed in order to perform their job more efficiently, i.e., the inner business goals. The second meeting was designed to encourage Risk and Clinical Managers to refine the business goals in such a way that they become measurable, thus obtaining the measurement goals. The result of these meetings is as follows:

Risk Management Group business goals:

- To identify and reduce risk to the hospital.
- To increase the commitment to patient satisfaction and service excellence.

Risk Management Group measurement goals:

- To develop a proactive approach to risk identification and mitigation through the provision of meaningful reports to the stakeholders who need it.
- To understand the current level of satisfaction of patients, staff, and visitors regarding the services provided by the hospital.

Internal Medicine Clinical Managers business goals:

- Staff needs to know the why, how, and when of events, mainly those that cause harm to patients.
- To know whether our unit is doing well or bad.
- To know if events tend to go up or down during some time interval.

Internal Medicine Clinical Managers measurement goals:

- To understand the cause, nature, and outcome of events occurring in the Internal Medicine unit in order to increase staff awareness.
- To understand where the clinical unit stands in comparison to sister units across departments and campuses.
- To understand the variations in number, nature, and outcome of incidents across the time dimension.

Step 5 – Pose questions

After the groups developed a set of goals capable of guiding the measurement exercise, the next step was to reach to a common understanding of what each goal really meant. In order to accomplish this, a meeting was scheduled with each group of stakeholders. During the planning phase, some time had been allocated to perform individual meetings with each stakeholder. This would have facilitated the collection of their ideas about the goals without external influence. A second meeting would have gathered all the stakeholders from one group, exposed all questions, and finally helped select the ones that better described the goal(s) at hand. Although this plan was a good idea, the reality at the hospital was that most personnel had very little time to spare. Therefore, the process had to take place in one single meeting per group, where all stakeholders participated.

The measurement manager, PSLS business analyst and PSLS project manager acted as mediators, guiding the discussion and encouraging the personnel involved in the project to express all of their ideas. The outcome of each meeting was a series of questions describing the goals.

It should be mentioned that, at this point, we discovered that the second goal of the Risk Management Group could not be achieved in this iteration, the main reason being the inexperience and unfamiliarity of the stakeholders with the section of PSLS that collected this particular information. It was decided that, for the moment, questions, and metrics would not be generated for that goal. It should also be emphasized that, although meetings took place independently, many of the developed questions were quite similar (further explained in the next section). See Appendix C for a comprehensive list of the questions developed.

An example of the developed questions by Risk Managers for the goal "*To develop a proactive approach to risk identification and mitigation through the provision of meaningful reports to the stakeholders who need it*" follows:

- What is the number of incidents reported by "Reporter"?
- What is the number of incidents reported by "Incident Type"?
- What is the number of incidents reported by "Location"?
- What is the number of incidents reported by "Date"? (Weekdays vs. Weekends)
- What is the number of incidents reported involving equipment? (Broken down per equipment?)
- What is the number of events reported involving medication? (Broken down per medication?)
- What is the number of events reported by "Notified personnel"?
- What is the number of incidents reported by "Incident Status"? (Per individual? Across a clinical unit?)
- What is the number of incidents reported per portfolio?

Step 6 – Develop metrics

By the end of September 2010, the Risk Management Group and the Internal Medicine Clinical Managers had developed a set of goals and their respective questions. The next planned step was to select criteria to help develop the metrics.

A meeting was scheduled with the project manager, the business analyst, the data analyst and the bio-statistician. This team selected the following basic criteria to be used as guidelines when developing the metrics:

1. It had to be considered whether the requested information was currently collected by PSLS. As PSLS is an extensive application, not all modules were actually implemented at TOH at the time, and no data was available for these absent modules.
2. The metrics and reports development had to be divided into two phases. Phase 1 defines and reports on metrics that can use the existing PSLS reporting infrastructure. Phase 2 will generate a different set of metrics (and respective reports) that require the develop-

ment of data models extracted from the organization's data warehouse, and likely the use of Cognos BI tools for reporting.

3. No developed metric should disclose patient personal information or ratings of individual events (as required by QCIPA). Therefore, all the information collected and reported should contain an appropriate level of aggregation.
4. To avoid interdepartmental issues and preserve confidentiality, each unit should be able to access data regarding their unit only. However, upper management would have access to all the information since they need to get the "bigger picture".

After defining criteria 3 and 4, the team realized that HC-GQM would have to address these issues in a direct manner and not as part of the guideline shown above. As stated in Section 1.3, a common issue faced by the health care system relates to concerns of confidentiality and disclosure of information, or more precisely the disclosure of patient information and of ratings of individual events. Furthermore, in doing this, doctors, nurses, and the organization itself, who may potentially be implicated in lawsuits, are also protected and will therefore be more cooperative and willing to participate in gathering and reporting on AEs and other incidents.

Consequently, the HC-GQM methodology needs to include one additional step to address these particular issues individually, thereby leading users to consider confidentiality and non-disclosure of information. This new step, which is inserted right before the development of the metrics, is called "Check compliance with privacy and confidentiality legislation". It analyzes the potential metrics to be developed and verifies whether the information that they provide complies with rules and regulations such as QCIPA. This means that any metric that reveals or can potentially reveal patient personal information or the details and ratings of a particular event should not be developed. This meeting among the project manager, the business analyst, the data analyst, and the bio-statistician, together with interventions from the director of Clinical Quality and Performance Management, made it clear that it was important to have these issues addressed as a separate step, as opposed to keeping them as part of the criteria discussed where stakeholders have the authority to decide whether or not to consider compliance issues when creating metrics.

A further result of this meeting was a process of categorization and standardization of the developed questions, so they would adopt a more concise format. This request was introduced by

the bio-statistician. It was mainly based on the fact that some questions seemed to be repeated across the many goals and that there were questions contained in other questions. An example of this situation follows:

- What is the number of events reported involving falls?
- What is the number of events reported by "Incident Type"?

Knowing that "falls" is a specific "Incident Type" value makes it clear that the second question includes the first one. Therefore, the metrics developed for the second question can also serve the purpose of answering the first one.

All of the questions were then analysed and summarized in such a way that the development of metrics could be performed in a simpler way. In some cases, this task was carried with relative ease. This is the case for questions posed by the Clinical Managers to address their first two goals. Questions posed for goal "To understand the cause, nature, and outcome of incidents occurring in the Internal Medicine unit in order to increase staff awareness" were basically containing those posed for goal "To understand where the clinical units stand in comparison to sister units across departments and campuses". An example follows for better understanding:

- The first goal proposes: *What are the number and percentage of events reported involving falls?* This question has to count the number of events that were reported as falls for a specific internal medicine unit.
- The second goal proposes: *What is the average number of events reported involving falls among the Internal Medicine units of TOH*? In this case, the number of events reported as falls also has to be counted. The only difference is the location where the events happen and the final processing of that information in order to develop the metric.

After analyzing each goal and their questions for both groups of stakeholders, the following derived questions were developed:

Q1. What are the number of incidents and respective percentages reported by "Event Type" during a "Time Period", for a determined "Location"?

Q2. What are the number of incidents and respective percentages reported by "Incident Type (Classification level 1)" during a "Time Period", for a determined "Location"?

Q3. What are the number of incidents and respective percentages reported by "Incident Type (Classification level 2)" during a "Time Period", for a determined "Location"?

Q4. What are the number of incidents and respective percentages reported by "Incident Type (Classification level 3)" during a "Time Period", for a determined "Location"?

Q5. What is the number of incidents reported by "Notification Level (Was someone notified?)" during a "Time Period", for a determined "Location"?

Q6. What is the number of incidents reported by "Documentation level (Documented in chart)" during a "Time Period", for a determined "Location"?

Q7. What is the number of incidents reported by "Level of harm (harmed patient, non harmed patient)" during a "Time Period", for a determined "Location"?

Q8. What is the number of incidents reported by "Potential of harm" during a "Time Period", for a determined "Location"?

Q9. What is the number of incidents reported by "Cause of harm (Management or Patient's underlying disease)" during a "Time Period", for a determined "Location"?

Q10. What is the number of incidents reported by "likelihood of causing harm (likely, non likely)" during a "Time Period", for a determined "Location"?

Q11. What is the number of incidents reported by "Reporter" during a "Time Period", for a determined "Location"?

Q12. What is the number of incidents reported by "Incident Status (approved, being reviewed, awaiting final approval)" during a "Time Period", for a determined "Location"?

Q13. What is the time difference between incidents happen and get captured during a "Time Period", for a determined "Location"?

Q14. What is the ratio of events per patient during a "Time Period", for a determined "Location"?

Q15. What are the number of incidents and respective percentages reported by "Medication" during a "Time Period", for a determined "Location"?

Q16. What are the number of incidents and respective percentages reported by "Medication route" during a "Time Period", for a determined "Location"?

During meetings held in October 2010, where the measurement manager, the project manager, the business analyst, the data analyst, and the bio-statistician participated, a set of metrics was

developed. This process was divided into Section A and Section B. Section A was in charge of developing the metrics that could address the "categorized" questions. Section B carried the responsibility of formally defining them.

It should be mentioned that the outcome of Sections A and B only included those metrics that could be addressed during Phase 1. Table 8 shows an example of the resulting metrics for Phase 1.

Question	Metrics
Q1	- Count of specific Event Type during "time range" by "location" - Total count of Event Type during "time range" by "location" - Sum of count of Event Type during "time range" for all locations - Percentage of specific Event Type during "time range" by "location" over the total number of events
Q2	- Count of specific Incident Type during "time range" by "location" - Total count of Incident Type during "time range" by "location" - Sum of count of Incident Type during "time range" for all locations - Percentage of specific Incident Type during "time range" by "location" over the total number of incidents
Q3	- Count of specific "level of classification 2 events" during "time range" by "location" - Total count of "level of classification 2 events" during "time range" by "location" - Sum of count of "level of classification 2 events" during "time range" for all locations - Percentage of specific "level of classification 2 events" during "time range" by "location" over the total number of events
Q4	- Count of specific "level of classification 3 events" during "time range" by "location" - Total count of "level of classification 3 events" during "time range" by "location" - Sum of count of "level of classification 3 events" during "time range" for all locations - Percentage of specific "level of classification 3 events" during "time range" by "location" over the total number of events
Q5	- Ratio of specific Notification Level during "time range" by "location" over the total number of patient safety events
Q6	- Ratio of events "Documented in chart" during "time range" by "location" over the total number of patient safety events
Q7	- Count of specific "Level of Harm" during "time range" by "location" - Total count of "Level of Harm" during "time range" by "location" - Sum of count of "Level of Harm" during "time range" for all locations - Percentage of specific "Level of Harm" during "time range" by "location" over the total number of patient safety events
Q8	- Count of specific "Harm potential" during "time range" by "location" - Total count of "Harm potential" during "time range" by "location" - Sum of count of "Harm potential" during "time range" for all locations - Percentage of specific "Harm potential" during "time range" by "location" over the total number of patient safety events
Q9	- Count of specific "Cause of harm" during "time range" by "location" - Total count of "Cause of harm" during "time range" by "location" - Sum of count of "Cause of harm" during "time range" for all locations

	- Percentage of specific “Cause of harm” during “time range” by “location” over the total number of patient safety events classified as harmful
Q10	- Count of specific “Likelihood to cause harm” during “time range” by “location” - Total count of “Likelihood to cause harm” during “time range” by “location” - Sum of count of “Likelihood to cause harm” during “time range” for all locations - Percentage of specific “Likelihood to cause harm” during “time range” by “location” over the total number of events classified as non harmful
Q11	- Count of specific “Reported by” during “time range” by “location” - Percentage of specific “Reported by” during “time range” by “location” over the total number of events
Q12	- Count of specific “Incident Status” during “time range” by “location” - Total count of “Incident Status” during “time range” by “location” - Sum of count of “Incident Status” during “time range” for all locations - Percentage of specific “Incident Status” during “time range” by “location” over the total number of incidents
Q13	- Median of time between when an incident happens and when it gets entered in the system
Q14	- Ratio of incidents per patient
Q15	- Count of specific “Medication” during “time range” by “location” - Total count of “Medication” during “time range” by “location” - Sum of count of “Medication” during “time range” for all locations - Percentage of specific “Medication” during “time range” by “location” over the total number of patient safety events
Q16	- Count of specific “Medication route” during “time range” by “location” - Total count of “Medication route” during “time range” by “location” - Sum of count of “Medication route” during “time range” for all locations - Percentage of specific “Medication route” during “time range” by “location” over the total number of patient safety events

Table 8 Metrics Developed for Phase 1 During Section A

Section B handled the formal definition of the metrics. Such definitions were not as thorough as expected since the team was using the PSLS infrastructure to develop the reports (using a graphical user interface). This meant that no formal definitions were necessary unless there were difficulties understanding a given metric. The team members in charge of this activity were the business analyst and the data analyst. They were provided with samples of formal definitions in case they had to use them. Such samples were developed using an SQL-like query language. Table 9 shows some of the definitions generated, as examples.

Metrics	Formal Definitions
Total count of "Medication" during "time range" by "location"	Count of "recordId" where "show_medication" = true and "inc_dincident" between Date1 and Date 2 and "tcs" = Location1
Sum of count of "Likelihood to cause harm" during "time range" for all locations	Count of "recordId" where "inc_n_nearmiss" = true and "tcs" = all and "inc_dincident" between Date1 and Date 2

Table 9 Examples of Formal Definitions of Metrics

5.2.3 Reports and Analysis Phase

The reports and analysis phase started in October 2010, after the development of a set of goals and their respective questions. The process of building report prototypes was an exercise carried simultaneously with the development of metrics. These activities complemented each other in that the development of report mock-ups was necessary to think about the format reports would have, who their intended audience was, and more importantly what information (i.e., the metrics) reports would contain. On the other hand, the more metrics were discovered, the easier it was to think about a report structure that would best accommodate them. Figure 7 shows a sample mock-up for one of the report views.

Used Filters: **Location**: Riverside Campus, **Time Range**: August 2010, **Service**: All

Page 1

Event Type	Total per Category	Percent
Patient Safety	20	20%
Environmental Hazard	50	50%
Security	10	10%
Clinical Observation	20	20%
Total	**100**	

If Patient Safety event is clicked…

Approval Status	Total per Category	Percent
Being reviewed	8	*40%*
Awaiting final review	5	*25%*
Approved	7	*35%*
Total	**20**	

Note: The highlighted row is the option the user clicked.

Figure 7. Sample Report Mock-Up

In order to report on certain information by criteria like time or location, a set of *dimensions* were identified. These dimensions are helpful in providing a reporting system with more interactivity and effective ways (e.g., with filters or drill-up or drill-down exploration) of quickly providing the user with the information that he/she requires. For example, in the report mock-up presented in Figure 7 by making use of the "Location" dimension, only records related to the "Riverside Campus" are shown. In the same way, the "Time Range" dimension limits the report outputs to events that occurred in August 2010. The potential usage of dimensions goes beyond this kind of simple filtering. The hierarchical structure of dimensions can be used to drill-down into specific details. For example, in Figure 7, by clicking on Patient Safety (along the Event Type dimension), the details of the sub-categories (here Being reviewed, Awaiting final review, and Approved) can be displayed.

Table 10 describes the various dimensions resulting from the meetings.

Dimension	Description	Examples	Hierarchy (Top-to-Bottom)	Available
Time	When did the incident happen? (Time Range)	Year:2008 || 2010-05-06 to 2010-10-26	Year-Quarter-Month-Week-Day	Phase 1
Location	Where did the event happen? (Physical location)	General Campus, Inpatient ward/unit, 4 North-Psychiatry	Facility-Type of Location-Exact Location	Phase 1
Service	Responsible Service for Patient (cuts across campuses and physical locations)	Medicine || Emergency || Anaesthesia || Surgery	Service	Phase 1
Demographic	Patients' demographic information	0-18 years old, Male	Age-range, Gender	Phase 2

Table 10 Dimensions for the Reports

The availability of these dimensions was classified in either "Phase 1" or "Phase 2". This division represented a realistic approach as to when the dimension could be implemented. Phase 1 represented data that was already collected by the system and that was possible to report upon using the existing Datix-based PSLS infrastructure. However, Phase 2 required the use of a data warehouse to accommodate more sophisticated calculations using data that was not being cur-

rently stored in PSLS. Therefore, the dimensions under this phase would be addressed at a later time, once the reporting infrastructure would be moved from the simple Datix environment to the more complex and powerful Cognos environment.

In the same way, the metrics to be contained in each report were also analysed and classified according to the previous schema, i.e., for Phase 1 or Phase 2. This schema provided users with an idea of *which* reports would be accessible *when*.

The process of creating reports started by studying the common reporting needs for both groups of stakeholders. Some similar requirements were found. These mainly regarded events classified as "Patient Safety Events". Specifically, stakeholders seemed to be interested in their characteristics and in the harm caused to patients as a result of their occurrence. Such discovery allowed the team to develop a template capable of satisfying the needs of all stakeholders while reducing the number of substantially different reports to be constructed.

It was also noticed that some reports could follow a hierarchical structure. This meant that information regarding certain questions could be presented in the form of a hierarchy, where the most general information would be at the top and more specific data would reside at the very bottom. The report could be accessed from either end (top or bottom) depending on the user's needs. In this way one report could answer several questions at the same time.

The first eight report prototypes were generated by mid-October 2010. Three of those were hierarchical reports with drill-up and drill-down capabilities. The others were simpler reports but nevertheless insightful in providing the information that users requested. These eight report mock-ups were analysed by the team members. Some issues were found regarding the user interface. More specifically, the team was concerned about how "easily" stakeholders would understand the proposed reports. Since the prototypes were merely a proposition of data format (a table, a graph, etc.) with no actual data attached, they had to be easily readable by non-experts.

Changes were performed in order to make the prototypes more user-friendly. Some of these changes included the replacement of "possibly complex" metrics such as trends by *metrics place holders*. Since the prototyping process started simultaneously with the metrics development, the team did not exactly know yet which metrics would actually be generated. Therefore metrics place holders were incorporated to reserve a space in the reports for the next-to-come valid metrics.

The generation of prototypes ended once the metrics were properly developed. A document exposing the proposed reports was then distributed among the stakeholders (Clinical Managers, Risk Management group and the director of Clinical Quality and Performance Management) for their final approval. Once a positive confirmation was received from the users, the reports were taken to the development team in charge of the PSLS implementation at TOH. This team was then put in charge of building and later on deploying the resulting solution. At this time however, the reports are still unavailable to the end-users.

5.3. Evaluation of the Implementation Process

After the creation of metrics and of report prototypes, the case study had come to an end and the team was dissolved. The objectives defined for this exercise were met: a set of metrics and reports were designed to provide TOH personnel with the information they really needed and wanted, with the hope of helping to understand the causes and effects of AEs and incidents within the hospital, and eventually of guiding decision makers.

In order to further evaluate the usefulness of HC-GQM, it was also decided to make an online and anonymous survey to obtain, retrospectively, feedback from the stakeholders and the team. The first six questions target the six steps of HC-GQM, whereas the seventh and last question targets the methodology as a whole. Most of the questions were presented using a five-point Likert item (with a *Not Applicable* sixth option) to facilitate the process of filling out the survey and analysing its results. Additional open questions were included to allow participants to provide additional comments. It should be noted that this survey is strictly qualitative as the project involved only a few people, and therefore no statistical meaning can be derived from their answers. Yet, this survey allows us to gain insight on the usefulness of HC-GQM from the perspective of users and stakeholders.

This section briefly discusses each question, its objective, and its results. The results have been categorized as *Positive* (for responses that include "Strongly Agree" or "Somewhat Agree" ratings), *Negative* (for responses that include "Strongly Disagree" or "Somewhat Disagree" ratings), and Neutral (for responses that include "Neutral" or "Not Applicable" ratings). Out of the 9 participants involved in the implementation of HC-GQM for the PSLS (excluding the author of this thesis), 8 have answered the survey: 5 team members (in green in the following diagrams), two clinical managers (in blue), and one risk manager (in red).

Question 1: The objective of this question was to discover how useful is a team with specific roles and whether this is a key aspect that HC-GQM should keep. Table 11 shows the results for question 1.

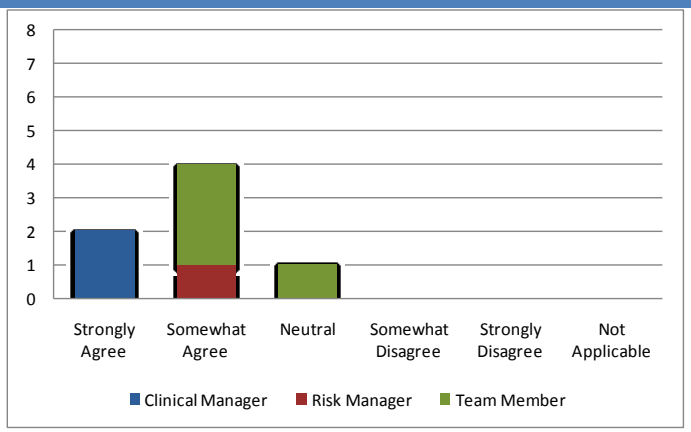
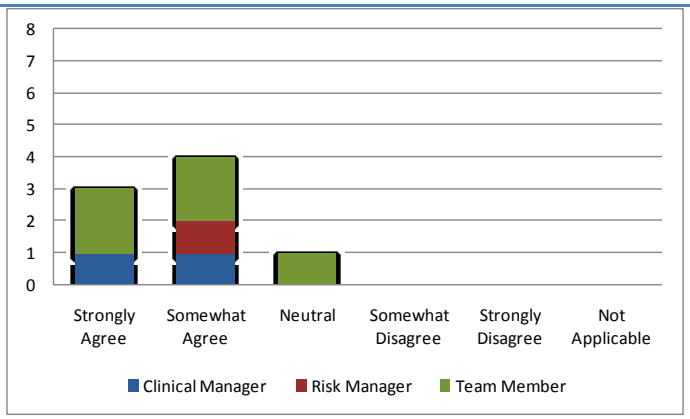
Question 1: The reports' requirement gathering was led by a team formed specifically for this purpose. Regarding the team aspects, please rate the following assertions	
Every team member had a specific role	
The team composition was a key aspect for the success of the project	

Table 11 Survey Results for Question 1

The results show that 7 out of 8 participants rated as *Positive* the fact that the composition of the team was a key aspect of success. Also, 7 out of 8 participants rated as *Positive* the fact that team members had specific roles within the project. No participant introduced *Negative* ratings. The clinical managers were especially positive for this question. One comment was provided to state that the business analyst should be the main figure when collecting the goals and questions. From these results, we can conclude that forming a team with the right members and stakeholders is an important step of HC-GQM.

Question 2: The objective of this question was to discover how useful an action plan is in order to keep stakeholders and team members informed on the project activities. Table 12 shows the results for question 2.

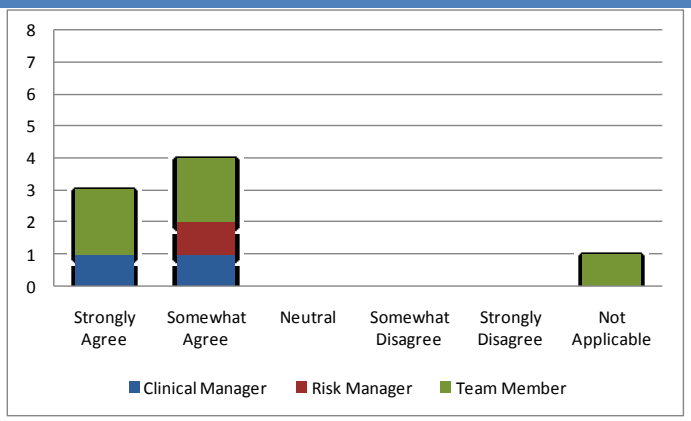
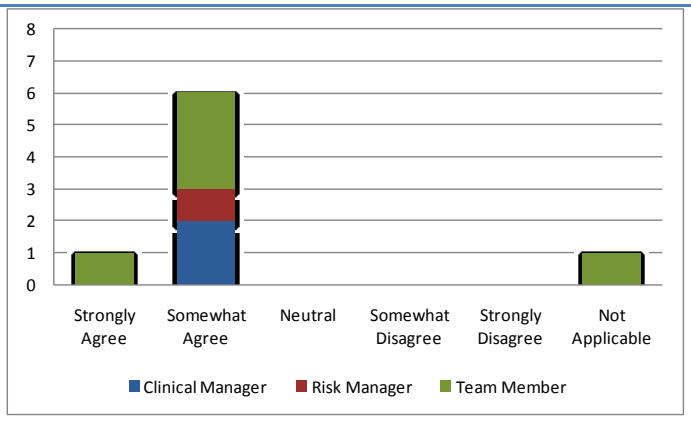
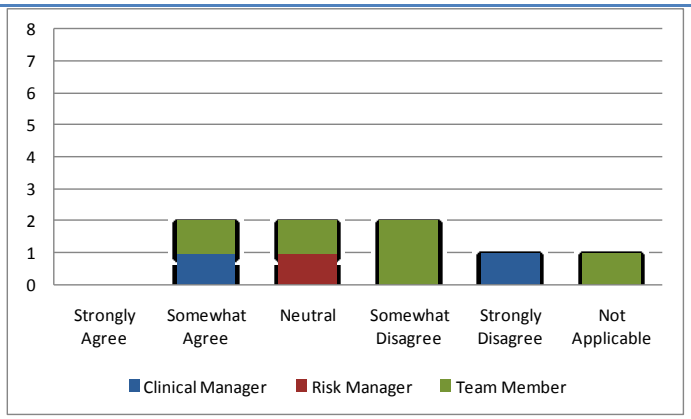
Question 2: To coordinate activities within the team and with stakeholders, an action plan was implemented. Regarding the plan aspects, please rate the following assertions:	
Such an action plan is necessary to keep all team members informed of future steps	
By looking at the action plan, I got an idea of the project's steps.	
The discussed topics were not properly addressed in the "meetings summaries" distributed after each meeting.	

Table 12 Survey Results for Question 2

The results show that 7 out of 8 participants rated as *Positive* the need to have an action plan in place. Also, 7 out of 8 participants rated as *Positive* the fact that the plan provided them with a road map to the activities to be accomplished. Answers regarding the meeting summaries are somewhat evenly distributed among all categories. This might indicate the need for improving the meeting summaries.

Question 3: The objective of this question was to discover whether the use of "entities" and the generation of goals had provided suitable guidance to the stakeholders in the process of finding metrics. Table 13 shows the results for question 3.

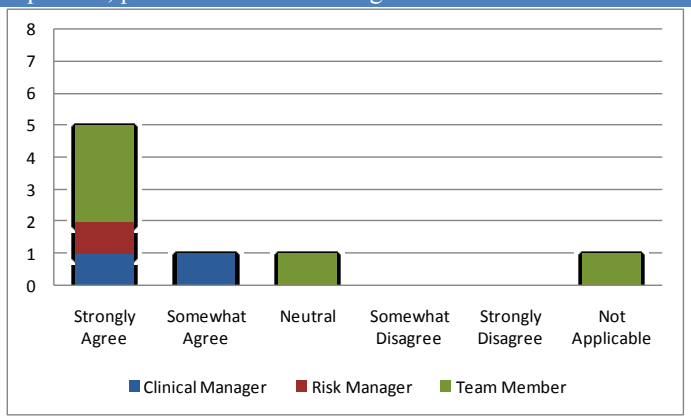
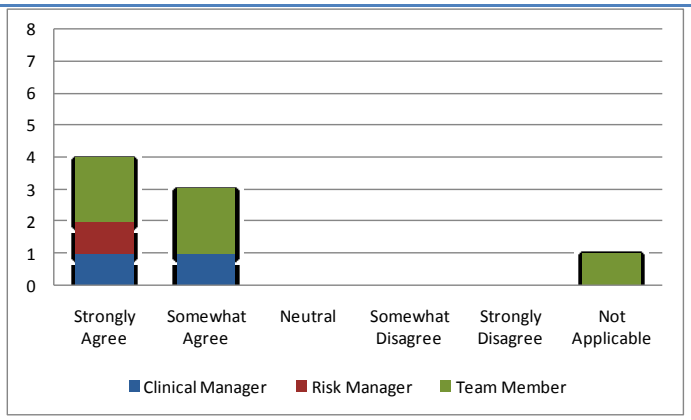
Question 3: In order to gather the requirements for PSLS reports, stakeholders were asked to take into consideration the process of collecting events and incidents. It was also asked to state "what" they were trying to achieve with these reports. Regarding this process, please rate the following assertions:	
By focusing on the PSLS processes of collecting "events" and "incidents", it was easy to think about possible indicators and reports.	
The creation of high-level "objectives" or "goals" helped me focus on the key issues that I was interested in.	

Table 13 Survey Results for Question 3

The results show that 6 out of 8 participants rated as *Positive* the use of PSLS's incidents and events as "entities" to guide the project. The other 2 participants had *Neutral* ratings. The use of goals was rated as *Positive* by 7 out of 8 participants. In this case, only 1 participant provided a

Neutral opinion. From these results, we can conclude that the generation of entities and goals provided the team and stakeholders with useful guidance.

Question 4: The objective of this question was to analyze the process of generating questions to describe the developed goal. Table 14 shows the results for question 4.

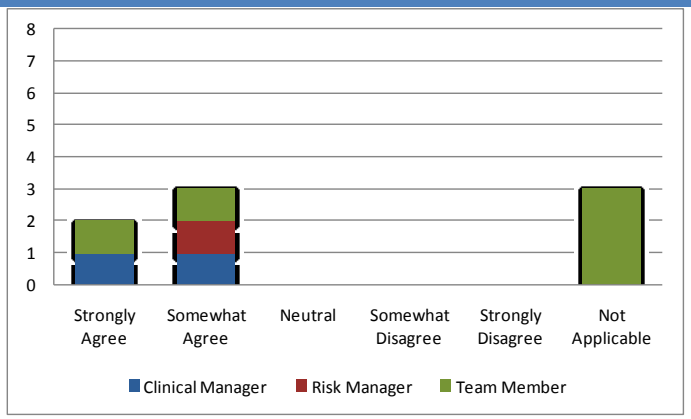
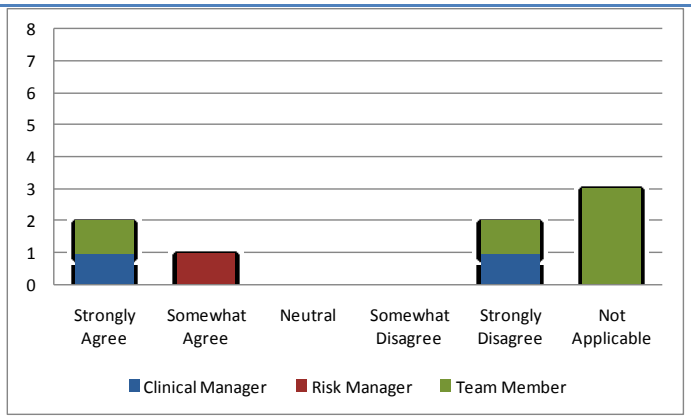
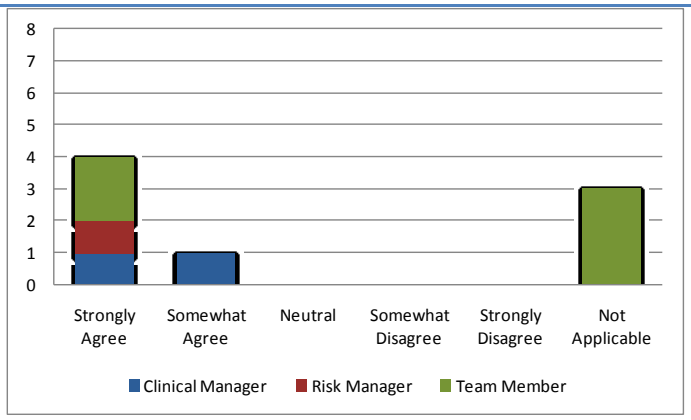
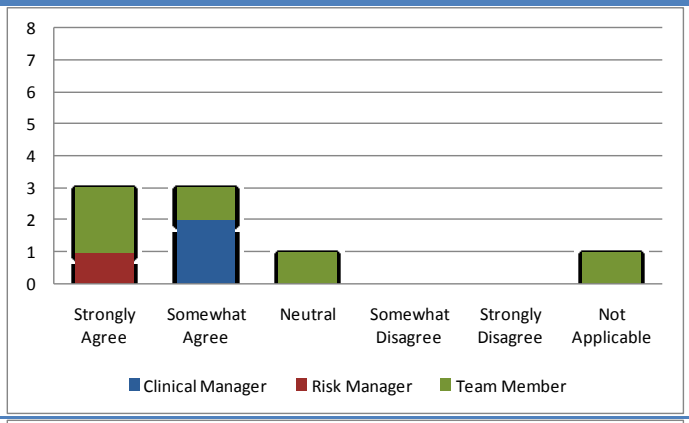
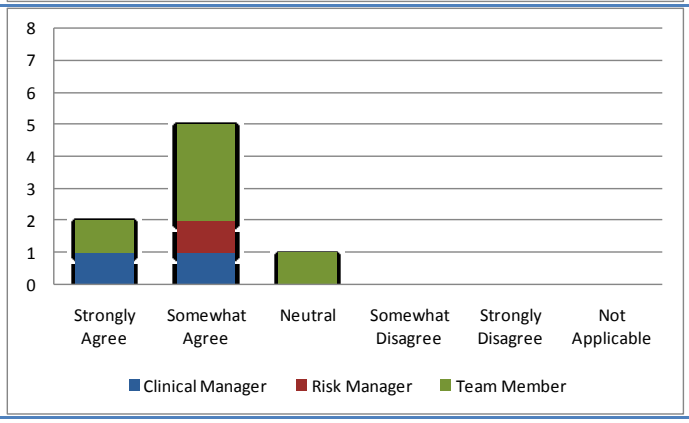
Question 4: After defining the goals, a series of questions were developed to describe them. Regarding this process, rate the following assertions:	
It was easy to use "questions" to describe the goals.	
Formulating questions in group made this task difficult.	
Posing questions was an easy way to visualize the information I wanted to report on.	

Table 14 Survey Results for Question 4

The results show that 5 out of 8 participants rated *Positive* the fact that it was easy to generate questions to describe the goals and to "visualize" the reports they wanted. There also were 3 *Neutral* responses and that corresponds to the number of participants who were not involved in that step of HC-GQM. The answers for the formulation of questions in group are divided among *Positive*, *Negative* and *Neutral*. This indicates that for some people, generating questions in group is easier than for others. This might represent a potential opportunity for improving the methodology in that different strategies (e.g., individually and/or in group) could be used to obtain the set of questions that stakeholders provide.

Question 5: The objective of this question was to discover whether the generation of certain "criteria" and "dimensions" helped towards the development of metrics. Table 15 shows the results.

Question 5: After the questions had undergone a categorization process, they were analyzed to see which indicators "or metrics" could be used to answer the questions. Regarding the metrics, please rate the following assertions: In my opinion:	
Generating criteria to know the availability of data helped me focus on the metrics that were possible to obtain.	
Thinking of the possible dimensions helped me generate the measures in an easy manner	

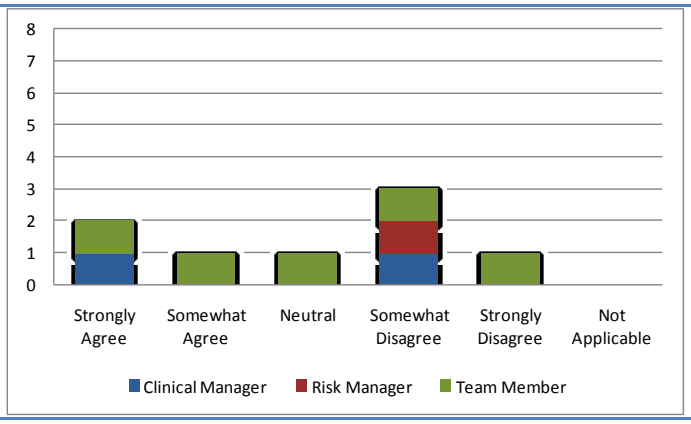
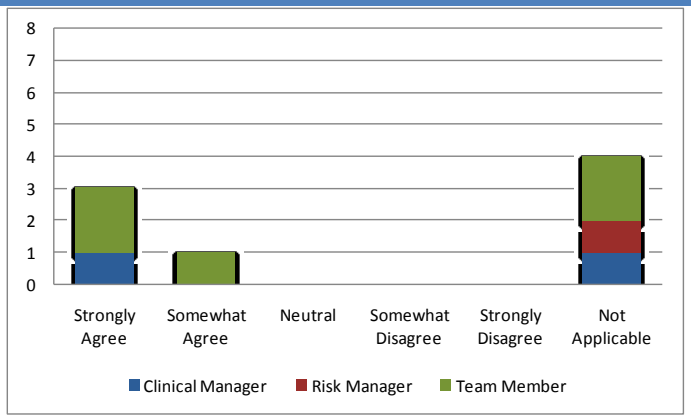
Generating indicators was not an easy task	

Table 15 Survey Results for Question 5

The results show that 6 out of 8 participants rated as *Positive* the generation of criteria to know the metrics that could be developed. There were 2 *Neutral* responses and that corresponds to the number of participants that were not involved in that step of HC-GQM. In the same way, the generation of dimensions was rated as *Positive* by 7 out 8 participants. In the case of the ratings related to how easy it was to develop the metrics, the answers were again divided mainly between *Negative* and *Positive*. *Negative* responses indicate that the task was considered as easy. From these results, we can conclude that team members and stakeholders have found useful the generation of criteria and dimensions to guide the development of metrics.

Question 6: The objective of this question was to analyze the process of generating prototypes of reports. Table 16 shows the results for question 6.

Question 6: A set of prototypes for reports was generated. Regarding this process rate the following assertions: In my opinion:	
Prototypes helped the team to develop the final reports.	

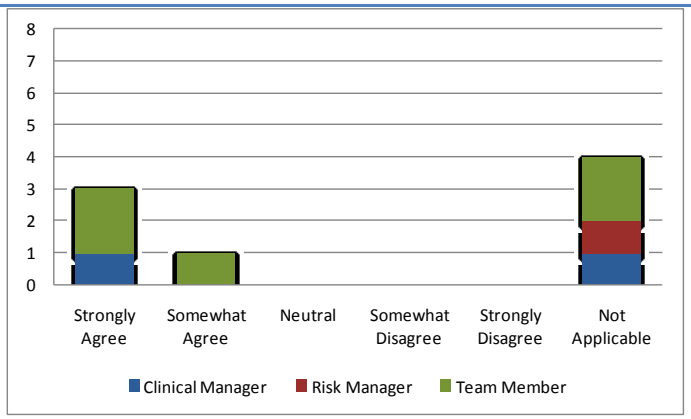
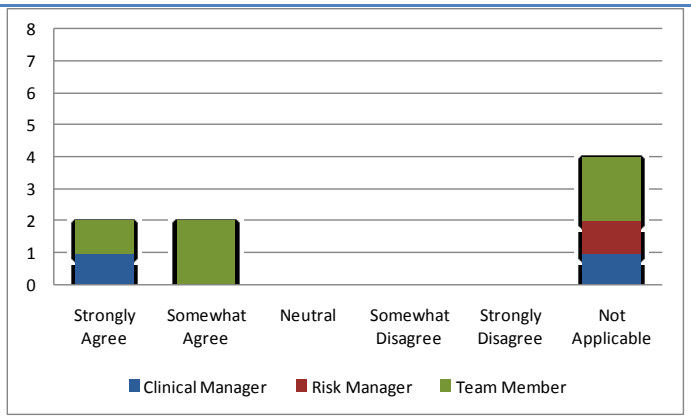
Prototypes were important to visualize needed indicators and reports.	
Prototypes represented the information stakeholders wanted from PSLS.	

Table 16 Survey Results for Question 6

The results show that 4 out of 8 participants rated as *Positive* the fact that prototypes facilitated the development of final reports as well as the visualization of the indicators that they should contain. In the case of ratings regarding whether prototypes represented the information stakeholders needed, the responses were evenly divided between *Positive* and *Neutral*. These answers imply that half of the participants involved in the survey did not receive the prototypes of the reports, an issue that should be resolved in future implementations.

Question 7: The objective of this question was to discover the general satisfaction of stakeholders and team members with the methodology and its outcomes. Table 17 shows the results for question 7.

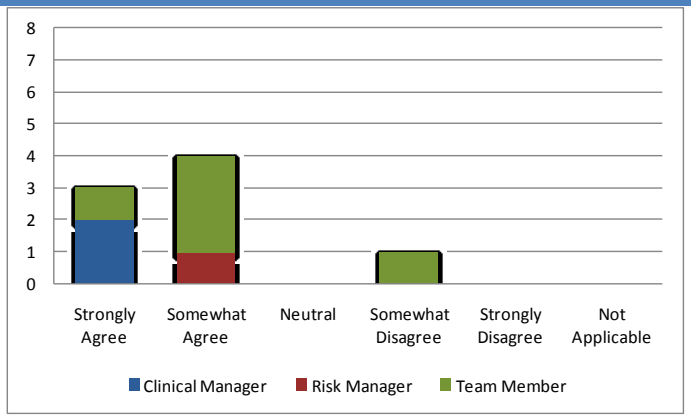
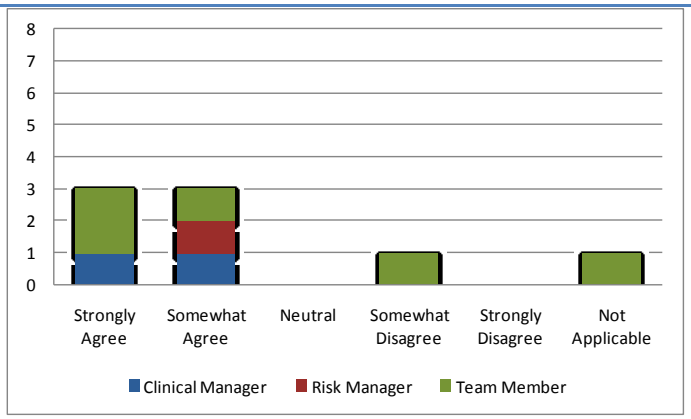
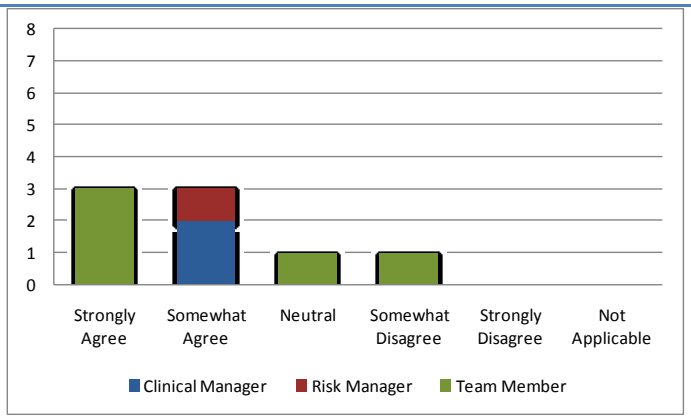
Question 7: How could you rate the overall process of defining indicators and reports? In my opinion:	
The experience has been satisfactory.	
The process steps were easy to accomplish.	
I would use this process again to obtain reports and indicators.	

Table 17 Survey Results for Question 7

The results show that 7 out of 8 participants rated the overall experience as *Positive*. Only 1 participant rated this experience as *Negative*. Regarding how easy the steps of the methodology were and the use of HC-GQM in the future, 6 participants out of 8 rated this aspect as *Positive*. The other two responses are 1 *Negative* and 1 *Neutral*. It should be pointed out that the respon-

dent who provided the *Negative* ratings for this question also rated all of the questions above as Neutral and was a team member. This leads us to believe that this person was a team member who was not very involved in the whole process. From these results we can conclude that most of the team members and stakeholders found the methodology easy to use and are willing to use it again.

5.4. Conclusion

This chapter focused on the implementation and validation of the refined HC-GQM at The Ottawa Hospital. The Patient Safety Learning System (PSLS) was described and introduced as a new tool for collecting incidents and adverse events. This application is also used to generate simple reports that do not require access to a data warehouse.

The six steps of the methodology were followed and in several instances modified to adapt to specificities of the PSLS application. Among the improvements added to HC-GQM are:

- To obtain an action plan approval from executive management in order to guarantee stakeholder involvement in the project. (Added to the Metrics Planning Phase, after the action plan is created.)
- To check that candidate metrics comply with privacy and confidentiality legislation, thus ensuring that patient personal information and individual ratings of events are not disclosed. (Added to the Metrics Development Phase, before the metrics are developed.)
- To categorize and standardize the questions in such a way that repeated questions are deleted, making the development of metrics smoother. (Added to the Metrics Development Phase, after the verification of the compliance of metrics against privacy and confidentiality legislation.)
- To create dimensions for the reports as a way of improving their effectiveness and making them more interactive. (Added to the Reports and Analysis Phase, before creating the report mock-ups.)

The chapter concludes by presenting and discussing the results obtained from a survey that was distributed among the team members and stakeholders to understand whether HC-GQM had been useful to them. Most of the results indicate a positive experience with the methodology.

Several points for improvement suggested by the results include a better usage and distribution of the meeting summaries, the generation of questions done individually and then in groups, and ensuring that the report prototypes are available to all members.

Chapter 6. Conclusions and Future Work

6.1. Conclusions

In this thesis, we have addressed a common problem faced by health care managers and clinicians: how to extract meaningful information from collected data. The solution proposed here is a goal-driven methodology for the development, collection, and analysis of meaningful metrics called *Health Care Goal Question Metric* (HC-GQM). HC-GQM adapts Basili's Goal Question Metric (GQM) approach, originally created for the software industry, to address issues specific to the health care sector. The outputs of HC-GQM are metrics and reports that pull data stored in databases or data warehouses and present relevant information to the interested stakeholders (e.g., doctors, researchers, and managers) in the format they understand and want, thus facilitating their decision-making process.

HC-GQM was implemented and tested using two separate health care quality contexts within the same teaching hospital. The first prototype implementation was done in the context of an Adverse Event Management System, to learn about the occurrences and consequences of adverse events in inpatient populations of the hospital. The results of this first study led to improvements to the methodology, which was then applied to the second context: an incident reporting system called Patient Safety Learning System. An online and anonymous survey showed that, the stakeholders and team members who participated in the implementation appreciated the methodology and are willing to use it again in the future. Furthermore, HC-GQM itself benefitted from this "double loop" validation process in that it steps became increasingly adjusted to the health care sector.

Through the use of HC-GQM, it was possible to elaborate patient safety reports that met the needs of health care stakeholders involved in both of the previously mentioned measurement exercises. It should be remarked that HC-GQM does not target any specific type of stakeholder. Rather, it benefits from having people from different backgrounds, with diverse interests and skills working together towards one common objective: the improvement of patient safety, and hence of health care quality.

In order to provide a better understanding on how HC-GQM is different to GQM, Table 18 is presented. This table summarizes the resulting HC-GQM steps, where they are defined, whether each step comes from GQM, as well as comments providing a brief rationale as to why steps where added, modified, or adapted from other related work.

HC-GQM Step	Defined in Section	GQM Step?	Comments
Metrics Planning Phase	4.3.2		This phase was added to handle project management issues.
Set the scope of the implementation.	4.3.2		A scope is added to know the scale of the implementation and therefore plan accordingly resources and time, as well as engage key stakeholders.
Create an action plan.	4.3.2		An action plan was added after the first implementation of HC-GQM as a way of organizing the work and, to get a perspective on the project steps, deadlines and time required to obtain the reports and metrics.
Get the action plan approved by executive management.	5.2.1		Executive management should decide on the implementation strategy and approve it. This also helps secure access to busy stakeholders.
Metrics Development Phase	4.3.1		This phase contains all the steps related to the development of metrics.
Form the team	4.3.1		This was not explicitly stated as a GQM step, but it is necessary for a successful implementation.
Select entities.	4.3.1		Selecting entities focuses the implementation on specific processes, areas, systems. This idea was taken from Fenton & Pfleeger [1997].
Select business goals.	4.3.1	X	It was recommended to perform a literature review to facilitate this process.
Develop measurement goals from the business goals.	4.3.1	X	No major changes were introduced.
Pose questions for the measurement goals.	4.3.1	X	Some aspects were added that GQM does not address: questions can be divided in sub-questions for better understanding; also certain questions might provide insights to define new goals (iterative aspect).
Check compliance with privacy and confidentiality legislation.	5.2.2		Step added to address privacy and confidentiality concerns.
Develop guidelines for metrics development.	4.3.1		Certain criteria can be defined to prioritize the development of some metrics over others thus tackling the generation of too many metrics (a GQM issue) [Berender et al., 2006]
Categorize the questions.	5.2.2		This step was added as a result of the implementation of HC-GQM in the context of PSLS. It facilitates the development of metrics.
Develop metrics to answer the categorized questions.	4.3.1	X	The development of metrics is supported by the generation of prototypes of reports.
Write formal definitions of the metrics.	4.3.1		Step added to avoid confusion and to clarify what the metric is really measuring.
Reports Generation and Analysis Phase	4.3.3		This phase was added to be able to run in parallel the development of metrics and the generation of report prototypes.

Develop tentative dimensions.	5.2.3		Step added to enhance the interactivity of the reports.
Create prototypes for the dimensional reports.	4.3.3		The creation of prototypes is important in that it allows the stakeholders to visualize the outcome of their work.
Obtain feedback and approval for the prototypes.	5.2.3		Step added to understand how satisfied stakeholders are with the results and to communicate possible requests.
Refine the prototypes.	5.2.3		Step added to enhance the reports prototypes when needed.
Develop the reports.	4.3.3	X	The reports will be the tool used to collect the metrics and obtain feedback on whether they address the goals or not. GQM does not explicitly make use of reports for this.
Analyze the reports.	4.3.3	X	Analyze if the metrics truly address the goals

Table 18 Comparison of Methodologies

It should be mentioned that in our implementation of HC-GQM, we did not analyse the resulting reports and determine whether the stated goals were addresses due to various delays within the hospital. This important step is something to look into more closely in the future, as this analysis would close the loop and validate that the results are what stakeholders need in order to improve their work.

Since this thesis describes each step of HC-GQM and provides examples of its implementation at a teaching hospital, it can also be used as a road map for health care institutions that consider applying this goal-based methodology to obtain tailored metrics and reports that are aligned with their goals.

The solutions and value provided by HC-GQM go beyond an adaptation of GQM to fit the specificities of health care. This methodology directly contributes to solving many of the difficulties that other approaches developed for the health care sector have been struggling with. Consequently, Table 19 recalls the problems stated in Section 1.3 of this thesis and provides brief explanations on how HC-GQM addresses each one.

Problem	Solution
Use of "one size fits all" metrics that are utilized in different contexts without taking into consideration their validity.	HC-GQM does not use the same metrics in all projects but it promotes the discovery of new metrics that fit the organization and current project's needs.
Metrics are generated following recommendations of "best practices" or through benchmarking techniques	Instead, HC-GQM proposes to find and generate the metrics through the development of goals. In this way the resulting metrics are in line with the projects' goals.
Concerns about confidentiality and non-disclosure of information that are linked to care providers being afraid that information that is not "confidential" can be used against them.	HC-GQM develops the metrics taking into consideration criteria from many different types of stakeholders, so a more heterogeneous view is available. Constraints related to privacy and confidentiality are also handled.
Some metrics are valid for some patient populations while the same set is not valid for others.	Reports and metrics are generated taking into consideration the point of view of the stakeholders defining the goals. Therefore, stakeholders from different areas will develop totally different metrics.
The fear of care providers to ruin their reputations or be engaged in lawsuits is another concern.	The methodology does not promote finger pointing, on the contrary, it considers the points of view of different stakeholders so the questions posed can describe the goals as thoroughly as possible.
Lack of clarity in who is responsible for the performance exercise.	By defining a team, stakeholders, and entities to be measured, HC-GQM makes a clear separation of roles and activities.

Table 19 Problems Solved

6.2. Threats to Validity

HC-GQM was validated by applying this methodology to the PSLS case study at TOH and by using a survey to collect the opinions of the stakeholders who used it. This section outlines challenges or threats that can affect the validity of the results presented in this research. They can be categorized as external or internal. External threats are those introduced by outside factors while internal ones are those relating directly to the validation process used.

One external threat is related to the composition of the team selected to implement the methodology. The team was formed primarily by personnel who had not been working for the hospital for more than a year. Therefore, their experience developing metrics and reports for the hospital was limited. This is considered a threat because without knowledge of other mechanisms to develop metrics, the team was lacking a comparison tool against which they could objectively analyze the steps and their outcomes. Nonetheless, this inexperience can also be regarded as a positive component, since their limited exposure to other methodologies allowed for an unbiased and fresh perspective on this work.

This threat was partially mitigated by running the resulting metrics and report prototypes by more senior staff at the hospital, since they possessed the necessary experience to advise on

this work. As an example, the outcome of the development steps for metrics and reports was validated by the director of Clinical Quality and Performance Management of the hospital.

Another external threat relates to the resulting metrics. The outcome of the implementation of HC-GQM at TOH was a set of fairly simple metrics (sums, counts, and percentages) and did not include complex metrics such as trends or forecasting. This threat has been categorized as external because it is the result of a decision made by the team and not a direct outcome of the methodology. By avoiding the development of more complex metrics, the team was bypassing some of the concerns clearly stated by the stakeholders, therefore augmenting the risk of not satisfying their goals and expectations. This threat is however handled by the iterative nature of HC-GQM, where the process can be repeated as many times as necessary until all the metrics are developed.

One last external threat is that the reports proposed for PSLS are still not fully implemented by the PSLS team or used by the stakeholders. This somewhat constraints the claims one can do about the usefulness of the reports for helping the decision-making process.

One internal threat that might affect the validity of the results relates to the tool used to obtain feedback after the implementation of the methodology, i.e., the online and anonymous survey. This survey was meant to extract information from the team and stakeholders involved. Although an effective technique, this type of validation can also potentially create a threat in that the questions and answers exhumed are subject to the researcher's interpretation. Therefore, bias may have been introduced in the results. A way of mitigating this problem was to run both the questions created for the survey and its results by the researcher's co-supervisors, for their approval. Also, the survey was made in such a way that followed a qualitative approach, focusing only on the degree of satisfaction of the stakeholders and the team regarding each of the steps of the methodology.

6.3. Future Work

For future work, we suggest that the methodology, with its additional steps, be applied again within the PSLS context. By reapplying HC-GQM, stakeholders and the team will be able to obtain better results and venture into phase two, which was left unfinished by the end of this work. Due to their familiarity with the process, its steps, and outcomes, we believe that after a second

implementation, a set of metrics and reports will be easier to obtain. What follows are several recommendations for reapplying HC-GQM.

- The generated prototypes of reports should be presented in a simple format. They should contain images, appropriate formatting, and examples of how the information will be displayed in the real reports. It is recommended to use HTML-based prototypes that users can access electronically, thus providing a "look and feel" closer to the real reports. These prototypes should be distributed among all stakeholders and team members for their consideration and feedback.
- Each and every member of the team should be taken into consideration, and not only the stakeholders. By communicating the project objectives, steps, and outcomes among all participants, everyone will be informed of the progress of the implementation. This also means that after each meeting, a summary of the meeting notes should be distributed to everyone on the team, and not only to the participants of the meeting.
- The process of collectively posing questions is not necessarily recommended. It is suggested that, if possible, this process should first happen individually. Then, all the questions gathered individually should be collected in a document to be distributed among the participants. After, a collective meeting can be held to reach a final agreement.
- It is also recommended to drive the meetings by sharing examples of goals and questions, as suggestions for the stakeholders set with the task to come up with their own goals and questions for their particular project. By providing this guidance and leadership, stakeholders will better understand the task at hand, thereby making the meetings more dynamic and productive.
- It should be taken into consideration as well the criteria dictated by Accreditation Canada to examine quality as part of the accreditation process. This could offer a pool of goals and questions to be considered by the stakeholders.

One way to further assess the generality of the methodology would be to apply it to the development of health care quality and performance metrics in an area different from patient safety or also in a different health care organization (e.g,, a different teaching hospital, a community hospital, or a medical clinic). In addition, it would be interesting to assess whether decisions taken by managers are actually influenced by the reports produced through HC-GQM. Lastly, HC-

GQM could benefit from further formalization, for example, by integrating some of the ideas presented by Kim et al. [2007], who introduce and use a measurement ontology based on the Toronto Virtual Enterprise (TOVE). Their approach helps defining a set of questions about behavioural and structural competencies that can then be formally expressed and measured against an axiomatic description of the organization.

References

Agency for Health Research and Quality – AHRQ (2010). PSNet Patient Safety Network. Glossary, published by the United States Department of Health and Human Services, retrieved December 2010, available at http://psnet.ahrq.gov/glossary.aspx#P.

Amyot, D., A. Forster, K. El Emam, L. Peyton, R. Tamblyn, and M. Weiss (2008). Performance Management at the Point of Care: Secure Data Delivery to Drive Clinical Decision Making Processes for Hospital Quality Control. NSERC/CIHR Collaborative Health Research Project, Canada, 2008-2011.

Arah, O. A., N.S. Klazinga, D.M.J. Delnoij, A.H.A. Ten Asbroek, and T. Custers (2003). Conceptual frameworks for health systems performance: a quest for effectiveness, quality and improvement. International Journal for Quality in Health Care, 15 (15), 377-398.

Aspden, P., J. Corrigan, J. Wolcott, et al., (Eds.) (2004). Patient safety: achieving a new standard for care. Committee on Data Standards for Patient Safety. Board on Health Care Services. Institute of Medicine of the National Academies. Washington, DC: National Academies Press, ISBN: 0-309-09077-6.

Bach, J. (1994). The immaturity of CMM. American Programmer, 7(9), 13-18.

Bache, R. M. and M. D. Neil (1995). Introducing metrics into industry: a perspective on GQM. Software Quality Assurance and Measurement: A Worldwide Perspective, (pp. 59- 68), International Thomson Computer Press, ISBN-10: 1850321744

Baker, G.R., P. G. Norton, V. Flintoft, R. Blais, A. Brown, J. Cox, et al. (2004). The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. CMAJ, 170 (11), 1678-1686.

Ballard, C., D. Farrell, A. Gupta, C. Mazuela, S. Vohnik (2006). Dimensional Modeling: In a Business Intelligence Environment. IBM Red Books, ISNB: 0738496448.

Basili, V. R (2005). Using Measurement to Build Core Competencies in Software. Seminar sponsored by Data and Analysis Center for Software.

Basili, V. R. (1993). Applying the Goal/Question/Metric Paradigm in the Experience Factory. Tenth Annual CSR (Centre for Software Reliability) Workshop, Application of Software Metrics and Quality Assurance in Industry, Amsterdam, Holland.

Basili,V. R., G. Caldiera, H. Rombach (1994). The Goal Question Metric Approach (pp. 528-532). Encyclopedia of Software Engineering, John Wiley & Sons, ISBN: 1-54004-8

Behnam, S., D. Amyot, A. Forsters, L. Peyton and A. Shamsaei (2009). Goal-Driven Development of a Patient Surveillance Application for improving Patient Safety. E-Technologies: Innovation in an Open World. LNBIP 26 (2), Springer, 65-76.

Berender, P. and P. Jönsson (2006). A Goal Question Metric Based Approach for Efficient Measurement Framework Definition. Proc. of the 2006 ACM/IEEE int. symposium on empirical software engineering. ACM, New York, USA, 316-325.

Blais, J., R. D'Angelo, M. Garzòn, P.M. Baillot and R. Bougueng (2009). Adverse Event Management System (AEMS). Final Report, Software Engineering Capstone Project, SITE, University of Ottawa.

Card, D. (2003). Integrating Practical Software Measurement and the Balanced Scorecard (p. 362). Computer Software and Applications Conference, Annual International, ISBN: 0-7695-2020-0

Card, D. (2003). Practical software measurement. Proc. of the 25th Int. Conf. on Software Engineering (ICSE '03). IEEE Computer Society, Washington, DC, USA, 738-739.

Curley, M. (2006). An IT value based capability maturity framework. Centre for Information Systems Research. MIT Sloan CISR Research Briefing, 6 (2).

Datix (2010) Software for Patient Safety, Incident reporting software for adverse event management. Retrieved October 2010, available at http://www.datix.co.uk/products-and-services/modules/incident-reporting

Donaldson M. S. (Ed.) (1999). Measuring the Quality of Health Care. A Statement by the National Round Table on Health Care Quality. Division of Health Care Services. Institute of Medicine (IOM), Washington, DC: National Academies Press, ISBN: 0-309-06387-6

Fenton, N. E. (1994). Software Measurement: A necessary scientific basis. IEEE Transactions on Software Engineering, 20 (3), 199 - 206

Fenton, N. E. and M. Neil (2000). Software Metrics: Roadmap (pp. 357-370). Proc. of the Conference on The Future of Software Engineering. ACM, New York, ISBN: 1-58113-253-0

Fenton, N. E. and M. Neil. (1999). Software metrics: successes, failures and new directions. The Journal of Systems and Software, 47 (2-3), 149-157

Fenton, N. E. and S. L. Pfleeger (1997). Software Metrics: A Rigorous & Practical Approach. PWS Publishing Company, Second Edition, ISBN:0-534-95425-1

Forster, A. J., T.R. Asmis, H. D. Clark, G. Al Salied, C. C. Code, S. C. Caughey, K. Baker, J. Watter, J. Worthington, and C. van Walraven (2004). Ottawa Hospital Patient Safety Study: Incidence and timing of adverse events in patients admitted to a Canadian teaching hospital. CMAJ. 170 (8).

Gold Practices Website – DACS (2010). Software Acquisition Gold Practice. Goal-Question-Metric (GQM) Approach. Retrieved June 2010, available at https://www.goldpractices.com/practices/gqm.

Government of Ontario (2004), Quality of Care Information Protection Act. Retrieved Dec. 2010 from http://www.e-laws.gov.on.ca/html/statutes/english/elaws_statutes_04q03_e.htm

Greenberg, A. H. Angus, T. Sullivan, and A. D. Brown (2005). Development of a strategy-based system-level cancer care performance indicators in Ontario, Canada. International Journal for Quality in Health Care, 17 (2), 107-114.

Groene, O., N. Klazinga, V. Kazandjian, P. Lombrail, and P. Bartels (2008). The World Health Organization Performance Assessment Tool for Quality Improvement in Hospitals (PATH): An Analysis of the Pilot Implementation in 37 Hospitals. International Journal for Quality in Health Care, 20 (3), 155-161.

Gurwitz, J.H., T. S. Field, J. Avorn, D. McCormick, S. Jain, M. Eckler, M. Benser, A.C. Edmonston, and D.W. Bates (2000). Incidence and preventability of adverse drug events in nursing homes. American Journal of Medicine, 109(2), 87-94.

Ho, D. (2004). Business Process Maturity Model- A CMM-Based Business Process Reengineering Research. Center of excellence in Business Process Innovation, Stevens Institute of Technology, retrieved November 2010, available at http://www.strategies-for-managing-change.com/support-files/businessprocessmaturitymodel.pdf

HP (2010). HP Neoview Advantage: A data warehouse platform for the new generation of business intelligence. Retrieved November 2010, available at http://h20195.www2.hp.com/V2/GetPDF.aspx/4AA0-2732ENW.pdf

Hulley, S. B., J. N. Martin, and S. R. Cumming (2001). Planning the measurements: precision and accuracy in: Designing clinical research: An Epidemiologic Approach. Lippincott Williams & Wilkins, Philladelphia, ISBN: 0781782104

IBM (2009). Reporting with IBM Cognos 8 Business Intelligence - Enterprise reporting that works the way you do. White paper, retrieved May 2010, available at http://www.zdnetasia.com/whitepaper/reporting-with-ibm-cognos-8-business-intelligence-enterprise-reporting-that-works-the-way-you-do_wp-1129783.htm

IEEE (1990). IEEE Std 610.12-1990 - Standard Glossary of Software Engineering Terminology. IEEE Computer Society, E-ISBN: 0-7381-0391-8

Institute of Medicine - IOM (2001). Crossing the Quality Chasm: A New Health System for the 21st Century. Committee on Quality of Health Care in America. Washington, DC: National Academies Press, ISBN: 0-309-07280-8

International Organization for Standardization - ISO (2007). ISO/IEC 15939:2007 - Systems and software engineering -- Measurement process. Retrieved November 2010, available at http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=44344

International Quality Improvement Program - IQIP (2010). Informational Brochure. Retrieved July 2010, available at http://www.internationalqip.com

Kahan, B. and M. Goodstadt (1999). Continuous quality improvement and health promotion: Can CQI lead to better outcomes? Health Promotion International. Oxford University Press, 14(1), 83-91.

Kazandjian, V. A., K. Wicker, S. Ogunbo, and N. Silverman (2005). Understanding safer practices in health care: a prologue for the role of indicators. Journal of Evaluation in Clinical Practice, 11(2), 161-170.

Kazandjian, V. A., P. Wood, and J. Lawthers (1995). Balancing Science and Practice in Indicator Development: The Maryland Hospital Association Quality Indicator (QI) Project. International Journal for Quality in Health Care, 7 (1), 39-46.

Khamis, N., S. Idris, and R. Ahmad (2008). Applying GQM Approach towards Development of Criterion-Referenced Assessment Model for OO Programming Courses. International Journal of Human and Social Sciences, 3(5), 333-336.

Kim, H., A. Sengupta, M.S. Fox, and M.M. Dalkilic (2007), A Measurement Ontology Generalizable for Emerging Domain Applications on the Semantic Web. Journal of Database Management, 18(1), 20-42.

Kimball, R. and M. Ross (2002). The Data Warehouse Toolkit. The Complete Guide to Dimensional Modeling. Second Edition. Wiley Computer Publishing, ISBN:0-471-200224-7.

Kohn, L. T., J. M. Corrigan, and M.S. Donaldson (Eds.) (1999). To Err is Human: Building A Safer Health System. Institute of Medicine (IOM). Washington, DC: National Academies Press, ISBN: 0-309-06837-1

Kurian, R (2009). The Benefits to Management of using GQM, Continuous GQM and V-GQM in a Measurement Program. M.Sc. thesis, Kent State University, UK.

Lane, P. and V. Schupmann (2002). Oracle9i Data Warehousing Guide. Release 2 (9.2). Oracle Corporation.

LeBrun, M. (2009). Prospective AE Surveillance and Learning Model. A Key Component of a Comprehensive TOH Adverse Event Management System. Project plan for the implementation of an AE Management System. Internal document, The Ottawa Hospital.

Lohr, K. (Ed.) (1990). Medicare: a strategy for quality assurance. Vol. 1. Committee to Design a Strategy for Quality Review and Assurance in Medicare. Division of Health Care Services. Institute of Medicine (IOM). Washington, DC: National Academy Press, ISBN: 0-309-04230-5

McGarry, J., D. Card, C. Jones, B. Layman, E. Clark, J. Dean, and F. Hall (2002). Practical Software Measurement. Objective Information for Decision Makers. Addison-Wesley, ISBN: 0201715163

McLamb, J. T. and R. R. Huntley (1967). The hazards of hospitalisation. Southern Medical Journal, 60 (5), 469-472.

McMillian, L. (2007). Reporting and Dashboards with Cognos 8 Intelligence Business. Published by Dashboard Insight. Retrieved Nov. 2010, available at http://www.dashboardinsight.com/sponsored-articles/cognos-reporting-and-dashboards.aspx

Michel, P., J.L. Quenon, A.M. de Sarasqueta, and O. Scemama (2004). Comparison of three methods for estimating rates of adverse events and rates of preventable adverse events in acute care hospitals. British Medical Journal, 328(7433), 199-203.

Microsoft (2010). SQL Server Reporting Services (SSRS). Retrieved November 2010, available at http://technet.microsoft.com/en-us/library/ms159106.aspx

Mills, D. H (1978). Medical insurance feasibility study - a technical summary. Western Journal of Medicine, 128(4), 360-365.

Mouchakkaa P., and D. Danielewski (2010). Cognos 8 Architecture and Tuning. Chicago and Milwake Cognos User Group Meetings, Published by Cognos. Retrieved November 2010, available at http://www.brightstarpartners.com/documents/Cognos_8_Arch_and_tune.pdf

Negash,S. and P. Gray (2008). Business Intelligence. Handbook on Decision Support Systems 2. Springer, Berlin Heidelberg, 175-193.

Oracle (2010). Oracle Business Intelligence Suite Enterprise Edition Plus. Retrieved November 2010, available at http://www.oracle.com/us/industries/040114.pdf

Özkan. Y. and D. Onur. (2009). Adopting Software Quality Measures for Health Care Processes. Studies in Health Technology and Informatics. 2009;150, 394-398.

Phelps, B (2004). Smart Business Metrics: Measure what really counts and manage the difference. Financial Times Prentice Hall, ISBN: 0273663968

Practical Software and Systems Measurement - PSM (2004). Measures for Success - Lessons Learnt from Applying PSM. Retrieved September 2010, available at http://www.psmsc.com/Downloads/Other/IFPUG_PSM_Keynote%20-%20Sept%202004.pdf

Practical Software and Systems Measurement – PSM (2006). Methods of Operation. Retrieved May 2010, available at http://www.psmsc.com/Downloads/Other/PSMMethodsofOperations_Nov2006.pdf

Practical Software and Systems Measurement - PSM (2010). About PSM. Retrieved October 2010, available at http://www.psmsc.com/AboutPSM.asp.

Safety Moment (2010). Open Source Safety Management, Glossary. Retrieved December 2010, available at http://www.safetymoment.net/resources/glossary

SAP (2010). Business Information Data Warehouse. Retrieved November 2010, available at http://help.sap.com/saphelp_nw04/helpdata/en/b2/e50138fede083de10000009b38f8cf/content.htm

SAS (2010). SAS Business Intelligence. Retrieved November 2010, available at http://www.sas.com/technologies/bi/

Shull, F., C. Seaman, and M. Zelkowitz (2006). Victor R. Basili's Contributions to Software Quality. IEE Software, 23 (1), 16-18.

Spertus, J. A., K. A. Eagle, H. M. Krumholz, K. R. Mitchell, and S. T. Normand (2005). American College of Cardiology and American Heart Association Methodology for the Selection and Creation of Performance Measures for Quantifying the Quality of Cardiovascular Care. Journal of the American Heart Association, 111, 1703-1712.

Stead, W.W. and H. S. Lin (Eds.) (2009). Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions. Committee on Engaging the Computer Science Research Community in Health Care Informatics. National Research Council of the National Academies. Washington, DC: National Academies Press, ISBN: 0-309-13050-6

The Joint Commission (2010). Specifications Manual for National Hospital Inpatient Quality Measures. Topic Library Item. Retrieved Dec. 2010, available at http://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures

The Ottawa Hospital - TOH (2010). The Patient Safety Learning System. Internal documentation.

Thomas, E. J. and Petersen, L. (2003). Measuring Errors and Adverse Events in Health Care. Journal of General Internal Medicine, 18(1): 61-67.

Thomas, E. J., D. M. Studdert, and T. A. Brennan (2002). The reliability of medical record review for estimating adverse event rates. Annals of Internal Medicine, 136 (11), 812-816.

Thomson, R., S. Taber, J. Lally, V. Kazandjian (2004). UK Quality Indicator Project® (UK QIP) and the UK independent health care sector: a new development. International Journal for Quality in Health Care, 16(1), i51-i56.

Trum Phillip (2010). HOIM is where the health is. McGill Publications. Retrieved December 2010, available at http://publications.mcgill.ca/headway/2010/10/13/hoim-is-where-the-health-is

Van Solingen, R. and E. Berghout (1999). The Goal/Question/Metric Method: a practical guide for quality improvement of software development. McGraw-Hill Publishing Company. ISBN 0077095537

Veillard, J., F. Champagne, N. Klazinga, V. Kazandjian, O.A. Arah, and A. Guisset (2005). A performance assessment framework for hospitals: the WHO regional office for Europe PATH project. International Journal for Quality in Health Care, 17 (6): 487-496.

Walsh, K. (2000). Adverse events in Health Care: Issues in measurement. International Journal for Quality in Health Care, 9(1), 47-52.

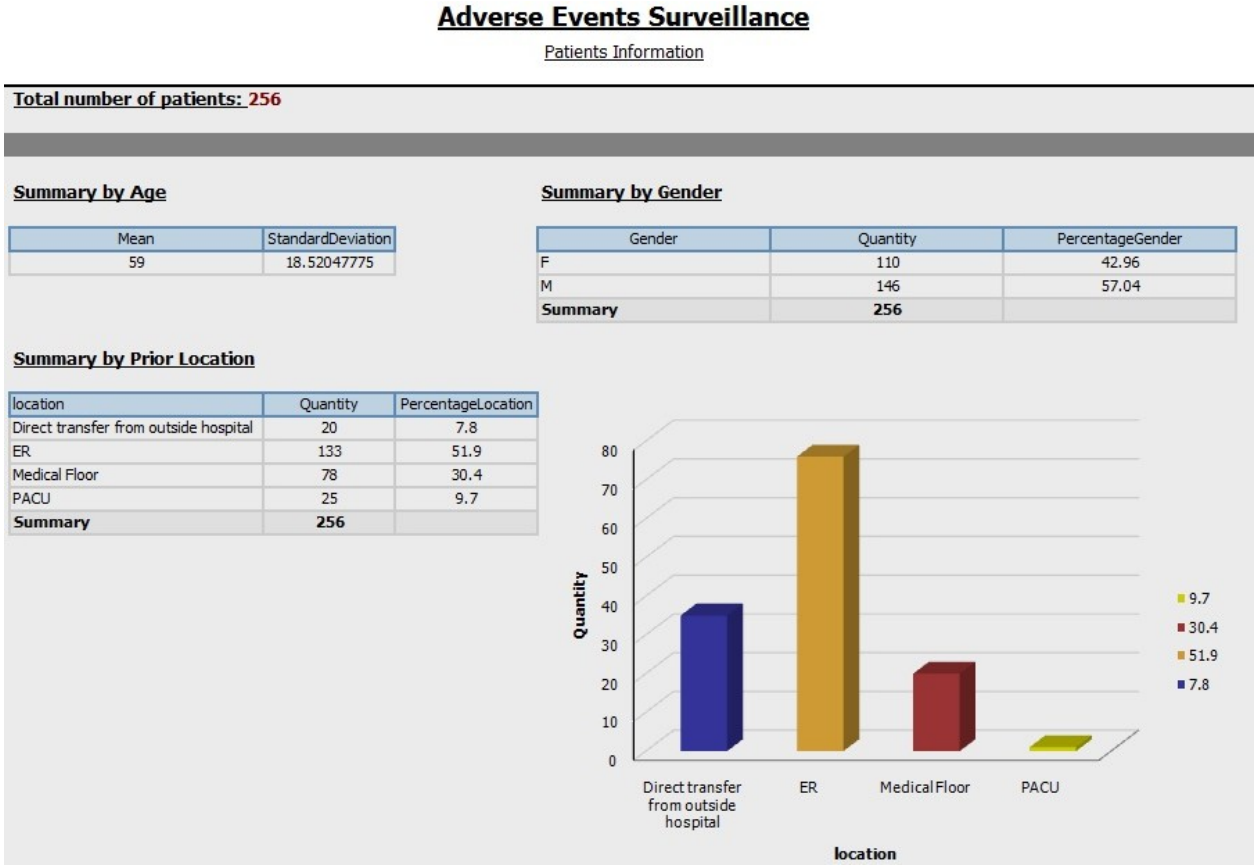
Walshe K (1998). The validity and reliability of adverse-event measures of quality in healthcare. Ph.D. thesis, University of Birmingham, UK.

Wikipedia, The free encyclopedia (2010). Data. Retrieved December 2010, available at http://en.wikipedia.org/wiki/Data

WHO EMRO (2004). Health Systems: Concepts, Design & Performance. Glossary. Retrieved Dec. 2010, available at http://www.emro.who.int/mei/mep/Healthsystemsglossary.htm

Appendix A - Examples of Reports for AEMS

Figure 8 shows an extract of the Patient Demographics report, produced with IBM Cognos 8 and populated with fake data.



Adverse Events Surveillance

Patients Information

Total number of patients: 256

Summary by Age

Mean	StandardDeviation
59	18.52047775

Summary by Gender

Gender	Quantity	PercentageGender
F	110	42.96
M	146	57.04
Summary	**256**	

Summary by Prior Location

location	Quantity	PercentageLocation
Direct transfer from outside hospital	20	7.8
ER	133	51.9
Medical Floor	78	30.4
PACU	25	9.7
Summary	**256**	

Figure 8. Patients Demographics Report Example

Figure 9 shows an extract of the Events report, also populated with fake data.

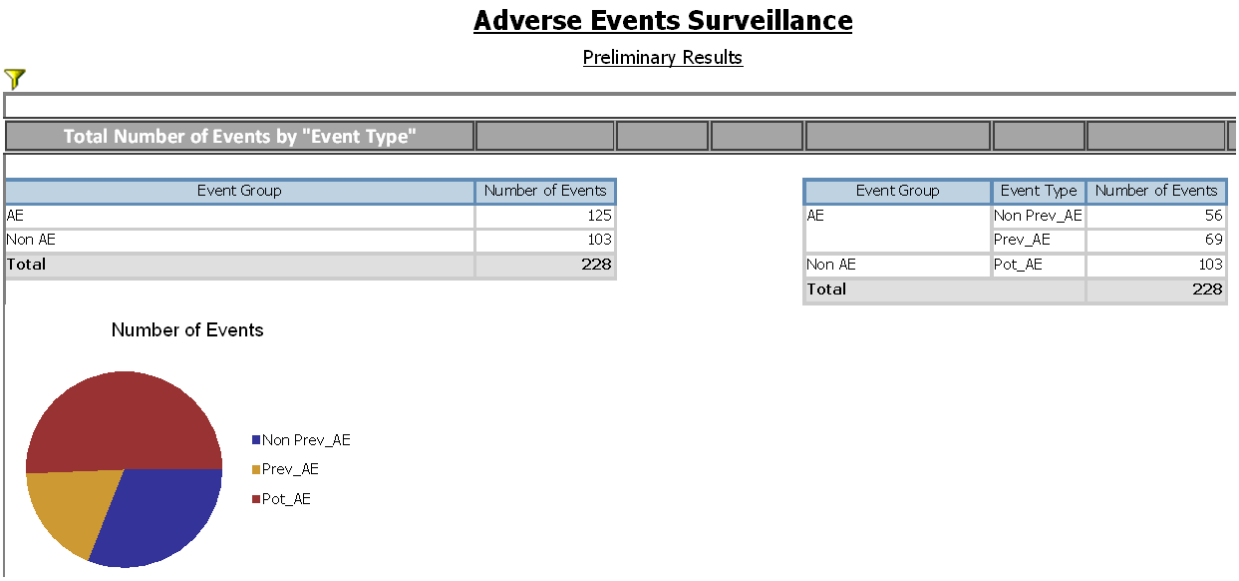


Figure 9. Events Report Example

Appendix B – Snapshot of Project Action Plan

Project Timeline

The following table represents the action plan timeline for the activities related to the second case study (PSLS, in Chapter 5). The different colours represent the three phases of HC-GQM. The italicized activities correspond to those activities that were not explicitly outlined in the methodology as steps but that were used during the implementation.

Activities to accomplish	Jun.20-Jul.2	Jul.2- Jul.10	Jul.10-Jul.12	Jul.12-Aug.3	Aug.3-Aug.14	Aug.14-Aug.30	Aug.31-Sep.22	Sep.22-Oct.1	Oct.1-Oct.7	Oct.7- Oct.15	Oct.15-Oct.30	Oct.30-Nov.5	Nov.5-Nov.26
Metrics Planning Phase													
Set project scope													
Select team members													
Create (or modify) action plan													
Initial plan approval													
Metrics Development Phase													
Select entities													
Select business goals													
Develop measurement goals													
Pose questions													
Develop metrics													
Generate criteria to follow for metrics development													

Activities to accomplish	June 20-July 2	July 2- July 10	July 10-July 12	July 12-Aug.3	Aug.3-Aug 14	Aug.14-Aug.30	Aug.31-Sep 22	Sep. 22-Oct. 1	Oct. 1-Oct. 7	Oct.7-Oct.15	Oct.15-Oct.30	Oct.30-Nov.5	Nov.5-Nov.26
Categorize questions													
Develop potential metrics according to criteria													
Formally define metrics													
Reports Generation and Analysis Phase													
Generate prototypes of reports													
Validate prototypes with stakeholders and team members													
Refine prototypes													
Generate survey and collect results													

Appendix C - Comprehensive List of Questions Generated by Stakeholders

The following is a list of questions generated by Risk Managers and Clinical Managers to describe their developed goals during the implementation of HC-GQM at TOH.

Risk Management questions for goal "*To develop a proactive approach to risk identification and mitigation through the provision of meaningful reports to the stakeholders who need it*":

- What is the number of incidents reported by "Reporter"?
- What is the number of incidents reported by "Incident Type"?
- What is the number of incidents reported by "Location"?
- What is the number of incidents reported by "Date"? (Weekdays vs. Weekends)
- What is the number of incidents reported involving equipment? (Broken down per equipment?)
- What is the number of events reported involving medication? (Broken down per medication?)
- What is the number of events reported by "Notified personnel"?
- What is the number of incidents reported by "Incident Status"? (Per individual? Across a clinical unit?)
- What is the time difference between incidents that happen and that get captured?
- What is the number of events reported by "Severity of Harm"
- What is the number of events reported as "Adverse Events"?
- What is the ratio of events per patient?
- What is the number of events reported by "Gender"?
- What is the number of events reported by "Age Group"?
- What is the number of incidents reported by "Level of Classification"? (Level 1, 2 or 3)
- What is the number of incidents reported per portfolio?

Clinical Managers questions for goal "*To understand the cause, nature, and outcome of incidents occurring in the Internal Medicine unit in order to increase staff awareness*":

- What is the number of events reported as "Patient Safety Events"?
- What are the number and percentage of events reported involving medication? (Broken down per medication)
- What are the number and percentage of events reported by medication route? (Oral, intravenous, etc.)
- What are the top 5 medications involved in events?
- What are the number and percentage of events reported involving falls?
- What is the number of events reported by "Incident Type"? (this question will also contain the incident types: falls and medication)
- Where do events most frequently happen? Number and percentage of events reported by "Location"
- When do most of the events happen? Number and percentage of events reported by "date and time" (Weekdays vs. Weekends) (Days vs. Night vs. Early mornings)
- Who is generally notified when an event happens? Who is least notified? Number of incidents reported by "Notified personnel"
- How many events are documented in the chart? Number and percentage of events documented in charts.
- What are the number and percentage of events reported by "Severity of Harm"
- What are the number and percentage of events reported as "Harmful"?
- What are the number and percentage of events reported as "Harmful" that were caused by "an error"?
- What section of the inpatient population presents most of events? Number and percentage of events reported by "Gender" and "Age Group"?

Clinical Managers questions for goal "*To understand where the clinical units stand in comparison to sister units across departments and campuses*":

- What is the average number of events reported as "Patient Safety Events" among the Internal Medicine units of TOH?

- What is the average number of events reported involving medication among the Internal Medicine units of TOH?
- What is the average number of events reported by medication route among the Internal Medicine units of TOH?
- What are the top 5 medications involved in events among the Internal Medicine units of TOH?
- What is the average number of events reported involving falls among the Internal Medicine units of TOH?
- What is the average number of events reported by "Severity of Harm" among the Internal Medicine units of TOH?
- What is the average number of events reported as "Harmful" among the Internal Medicine units of TOH?
- What is the average number of events reported as "Harmful" that were caused by "an error" among the Internal Medicine units of TOH?

Clinical Managers questions for goal "*To understand the variations in number, nature, and outcome of incidents across the time dimension*"

- What is the trend of incidents reported as "Patient Safety Events" for a given date range?
- What is the trend of incidents reported involving medication for a given date range?
- What is the trend of incidents reported by medication route for a given date range?
- What is the trend of incidents reported involving falls for a given date range?
- What is the trend of incidents reported by "Incident Type" for a given date range?
- Where do incidents most frequently happen? Trend of incidents reported by "Location" for a given date range.
- What is the trend of incidents reported by "Severity of Harm" for a given date range?
- What is the trend of incidents reported as "Harmful" for a given date range?
- What is the trend of incidents reported as "Harmful" that were caused by "an error" for a given date range?
- What is the trend regarding "Gender" and "Age Group" in the inpatient population for a given date range?