Research Paper

Patient Data Collection Methods. Retrospective Insights.

Julia Ferencz, MD[*1],Cosmin-Nicolae Serban[2]
[1]*Iuliu Hatieganu University Of Medicine And Pharmacy, Cluj-Napoca, Romania*
[2]*EMAC SRL, Cluj-Napoca, Romania*

ABSTRACT
Introduction: *Multiple classic and modern data collection techniques are presented in the current paper, but only a mix of them provides the appropriate approach to address patient safety problems. The current study aims to reveal the data collection methods applied worldwide.*
Materials and Methods: *All scientific sources of the current article were identified mainly by research on Internet. The matching words used in the search of materials are "data collection methods", "hospital reporting systems", "incident reporting systems", "patient events", "patient reported data". Relevant articles and studies covering the 2003-2016 timeframe were selected as a reference.*
Results: *Various data collection procedures are available worldwide. During several years of research, it was concluded that a significant number of patient studies use the following patient data collection methods: retrospective record review, record review of current inpatients, staff interview of current inpatients and nominal group technique based consensus method.*
Conclusion: *New trends in data collection techniques are also discussed, as they reveal the potential of the electronic environment. Future insights on this topic should consider the standardization of different data collection methods in order to improve data comparability aspects.*
Keywords: *EHR systems, health care, incident reporting, medical errors, patient data collection*

I.   INTRODUCTION

Nowadays, we have reached a high level of knowledge that empowers us to provide relevant expertise in several key domains. The technological evolution has offered us the possibility to constantly improve the quality of our life. Regarding the health care system, there was always a constant preoccupation with patient safety matters, for error prevention procedures. From this perspective, every medical unit and health care provider is searching for the perfect data collection system. In order to identify areas of improvement, the relevant role players must collect reliable patient data properly. Health care procedures and treatments have improved year on year. But, none of these would have been possible without intensive research, which represents the fundamental activity for taking the next step in our evolution. The advanced economies of the world constantly invest material, financial and human resources in medical research. Any deviation from standard medical procedures could result in patient injury.

The methods of data collection in hospitals depend on factors such as the final purpose of research, in terms of complexity and the amount of data, and on the available instruments of collection, in terms of research staff, specific protocols, tools and costs of the overall process. Other factors that influence these methods may be the geographical region, demographical configuration of individuals and their psychological status, hospital work environment or the availability of specific medical techniques at any given point in time. The current study aims to reveal the data collection methods applied worldwide.

II.   MATERIALS AND METHODS

All scientific sources of the current article were identified mainly by research on Internet. Electronic sources of data include the Wiley online library, the official web site of the U.S. Health Resources and Service Administration (aka HRSA), the website of the Agency for Healthcare Research and Quality in the U.S., PubMed Central® (PMC) archive of biomedical and life sciences journal literature (U.S. National Institutes of Health's National Library of Medicine), the American Hospital Association data viewer, the British Medical Journal website, the Canadian Association of Provincial Cancer Agencies website, Elsevier, Springer

International Publishing, the International Journal of Medical Informatics, World Health Organization guidelines for patient incident reporting, Minnesota Department of Health, St. Michael's Hospital, the Council of Health Care Accreditation of Southern Africa, Google search. The matching words used in the search of materials are "data collection methods", "hospital reporting systems", "incident reporting systems", "patient events", "patient reported data". Relevant articles and studies covering the 2003-2016 timeframe were selected as a reference. A variety of data referring to different procedures of extracting and managing patient information was collected.

III. RESULTS

The volume of data that is available for specific research is increasing exponentially. Various data collection procedures are available worldwide. During several years of research, it was concluded that a significant number of patient studies use the following patient data collection methods: retrospective record review, record review of current inpatients, staff interview of current inpatients and nominal group technique based consensus method. A brief description of each method clearly explains the mechanism.

Retrospective record review: the classic patient information gathering method which is based on previous stored medical records. It represents a two-step data collection procedure, at the beginning by the relevant patients and then the occurrence of side effects is assessed by a senior physician, together with the causal factors and a follow-up plan.

Record review of current inpatients: a more efficiently adapted version of the record review classic method, in terms of time consumption and identification of real-time health care problems. It assumes the analysis, for all hospital inpatients, of their medical history.

Staff interview of current inpatients: this method relies on interviews with health centre personnel. The superiority of this method can be harnessed in habitats with lack of relevant data or unavailable medical evidence.

Nominal group technique based consensus method: a recent data collection method which involves an assessment discussion of specific topics, based on contributors' personal insights. It requires no complex training, limited facilities, and offers fast implementation. Every participant opinion is very important. All ideas and solutions are noted on paper and at the end participants select the best solution.

Minnesota Department of Health indicates the following collection methods: focus groups, key-informant interviews, Community Forums and Public Hearings, surveys, Community Resource Inventories.

Focus Groups: a conference is organised for 5 to 10 people with comparable origins, education, experience, and the discussion is guided and recorded by a moderator. The highest-priority problems of the conference topics are discussed in depth. Even though the expertise required in this method is moderate and it involves low costs of implementation, it establishes the key issues and it provides new action points [1].

Key Informant Interviews: face-to-face interviews on the research problem with medical experts, who offer relevant answers about specific requirements regarding particular problems and the next steps to be followed. This method requires moderate expertise and it involves low costs of implementation. All data is recorded on paper and then is centralised for further analysis [1].

Community Forums and Public Hearings: public discussions with the community in order to assess and identify health care requirements. However, the costs and level of expertise are quite low. This method is inspired by political debates. The advantages of this method are: all individuals from the community can be part of the forum, it is inexpensive, can increase the community's perception about the problem and raises the opportunity to act at a management level. It stimulates community commitment to participate in the resolution plan [1].

Surveys represent complex instruments for gathering data, usually statistically significant data, with relevant level of expertise and a high cost of implementation. The main advantage of this approach is that it gathers a statistically significant amount of information from relevant targeted population.

Community Resource Inventories: this method represents a post-survey phase, when the summaries of medical service providers and their activities are investigated. It involves a medium cost of implementation, but it provides information about the service providers in a targeted area, the quality awareness of their services and their utilization rates, and future collaboration ventures between the community and local organizations [1].

In 2010 the World Health Organization published a guide for gathering patient data in hospitals in poor-developed economies and developing countries across the world. The guide is called "Assessing and tackling patient harm. A methodological guide for data-poor hospitals" and it was especially designed for these countries. It compiles the appropriate tools and applied methods for gathering patient data and it also describes the procedures to manage patient data, in order to reduce the level of harm from the medical system and to establish best practices with respect to patient protection measures. Methodological aspects of data collection refer to retrospective reviews of the hospital records, incident reporting data, and hospital audits [2].

Corresponding Author: Julia Ferencz, MD, email: j.ferencz@emac-consult.com

Matthew B. Weinger et all identified, in San Diego, USA, a variety of collection methods to report clinical events. It is well known that patient data is collected for various purposes such as quality improvement processes, for public or private use, or, of course, for research purposes. The authors of the "Retrospective data collection and analytical techniques for patient safety studies" article mention the difference between low-sample analysis procedures and industrial analysis procedures (e.g. cognitive task analysis interviews, failure-modes-and-effects analysis) [3]. The outcome of an individual analysis will explain clearly the factors that led to the specific patient injury. The statistics-based techniques applied on wide datasets offer a generalized preview of risk factors in the case of patient safety issues. Regardless of the purpose of the research, each method of data collection must highlight those patient injury abnormal effects [3].

When we refer to reporting structures, they should target providing valuable input figures that will serve the improvement of health care techniques, while ensuring all users on principles such as privacy, anonymity and data security. Based on participation criteria, these are grouped in two categories: compulsory reporting systems and voluntary reporting systems. Usually, compulsory systems are associated with public authorities and they lead to inaccurately reported events [3].

A proper example of a voluntary reporting system is the **Australian Incident Monitoring System**. This electronic system gathers significant amounts of data about medical events, which offer a comprehensive view of the factors and the necessary steps to remedy the situation [3].

A trend which has appeared recently in collecting patient information is the use of electronic devices and internet technology. These alternatives to the classic data collection methods ensure the anonymity and in the same time the quality of the reporting process. Specific reporting applications were created to facilitate the process [3].

In the "Retrospective data collection and analytical techniques for patient safety studies" paper the clinical supervision electronic system, which contains an index with all relevant elements regarding patient-related events: existing injuries, type of provided treatment, hospital parameters, is mentioned as a suitable data collection method. Independent entities such as professional organizations created the **"Medical Event Reporting System for Transfusion Medicine" (MERS-TM)** from Canada, which is able to smooth the way of reporting patient events, based on a clear and comprehensive form that can be completed by everyone. Non-compulsory association systems like **The Veterans Administration's RADARx,** the **Center for Disease Control's (CDC) VAERS,** and **United States Pharmacopeia's (USP) MedMARx** are other successful models of event reporting structures [3] [15].

The UK National Institute of Health Research supports medical research by the funding of **Health Technology Assessment (aka HTA) investigations**, in which methods such as patient questionnaires, the investigation of the available medical records or more advanced techniques, such as the use of a board of experts, are used. The purpose of HTA investigations is to assess the impact (from a medical and costs perspective) of health care systems, including tools, treatments, action plans and screening. All methods mentioned above have advantages and disadvantages. The use of daily patient data relies on proper recording of their data and appropriate, specific storage systems [4].

Several years ago, in Chicago, USA, a customized **electronic health record system** (aka EHR system) was created and implemented by the Alliance of Chicago Community Health Services. This system provides assistance in the decision-making process for clinicians and connects patient data with medical performance indicators. During 2005-2006 the centralized EHR system was applied at 32 clinical sites. This system was located in a locked space and the data could be retrieved through the Internet. Stored data revealed trends and comparisons between communities, between separate demographic categories. In 2007 in Massachusetts, US, all health care units were instructed to gather demographic patient data such as race and ethnicity, data which was integrated into the electronic system of every hospital. Data was sent to the US Division of Health Care Finance and Policy [5].

The most-utilized method of collecting patient data, at least in United States of America, remains the **survey administration**. This method is considered to constitute the central starting point for assessment of population health level and the requirements of the health care system, as surveys reveal non-administrative patient data. Not all surveys provide the expected outcome, only those which target health care plans, hospitals and health experts can be considered as a reference for health care amendment actions. Unfortunately, the majority of surveys represent a parallel data--gathering model as outcome data is not directly integrated in the health care system. Data reliability remains a key issue when we intend to use it in medical care upgrade action and medical error mitigation plans. Hospitals and health care units often collect significant amounts of relevant data, without using it properly for the improvement of the medical act [5].

In Toronto, Canada, the immigration process produced a significant modification of the city's demographics. As a consequence, a question arises: does the development of the medical system keep pace with the trend? The Ontario Human Rights Code and the Excellent Care for All Act (2010) acknowledged the focus

on patient care as a foundation for quality care. At the moment, in Canada, the health equity data is extracted from the **Canadian Community Health Survey** and/or related datasets [6].

A team of equity experts from three Canadian health care entities (Mount Sinai Hospital, Centre for Addiction and Mental Health and St. Michael's Hospital) agreed in 2009 to create an association that would fight for equity health treatment. The team of experts concluded that the health data collection structure needs to be upgraded and improved in order to achieve the association's purpose, as they found a lack of relevant data for the assessment of health care results at their establishment level. Later Toronto Public Health joined the action and brought a new perspective and a **new pilot scheme** was implemented. This pilot scheme accessed patient data from the following origins: the four above-mentioned health care organizations and data collectors. The health care organizations provided quantitative data through **patient studies**, while the second method of data collection refers to **focus groups, individual face-to-face interviews and online interviews** [6].

Even though the primary sources of data are considered to be hospitals, cancer centres, private doctor's offices, general specialists, screening programmes and central official records, the application of all these, at once, remains a visionary dream [7].

Regarding the sources of cancer identification, some methods involve the investigation of hospital patient records and extraction of those records for patients with cancer diagnosis. The reliability of this method is based on the accuracy of stored medical data and the correct recognition of cancer cases by the personnel. Another source of relevant data may be represented by outpatient clinics, as long as they manage to collect cancer case data. Even though the percentage of cancer cases detected in these health care units is small as compared to other sources of cancer identification, these cases must be included in the registers [7].

The most reliable cancer identification starting points are the pathology laboratories. These specific medical units identify, through a histological code, each disease that must be recorded. Uncertain cancer cases are also included in order to be assessed at a centralised level. When historical hospital registers are available, these registers should include copies of pathology reports. Duplicate data can be identified easily by using specific computer algorithms [7] [13].

After an individual's life has ended, the autopsy transcript is an essential piece of information. At this stage, all deaths caused by cancer should be registered with a different code. Haematology laboratories successfully detect leukaemia and lymphoma, while other types of medical investigation units, by using different diagnostic tests, identify different types of cancer. Screening programs have a long and successful history in cancer detection. Copies of radiotherapy results and observations, the case summaries, and hospital release forms represent an excellent method of data collection [7].

In general, we assist in the dispute between the manual data collection process and the electronic reporting systems that manage patient data. We can identify at least three general methods of collecting patient data: observational approach, backdated data mining from an electronic reporting environment, and investigation of inpatient health transcripts after the patient's hospital release. The choice of data collection method to be implemented is made by analysing the purpose of the research, the costs and the effect upon investigated patients. Prospective methods imply on-going observation for a clearly determined period of time [8].

In 2008 the American Hospital Association (aka AHA) launched an annual IT study amongst its members, in order to investigate the degree of digitalization within their organizations. As a result, an electronic health record system was implemented at state level, which represents the single point of relevant medical and quality information. Another survey was conducted by AHA in 2013. The goal of this second questionnaire was to assess and to observe the effects of previous EHR implemented systems. The survey covered more than 6,400 health care units across United States. Collected data refers to quality of health care services, sick person's fulfilment, and mortality. Public authorities refer to this data when they elaborate action plans with respect to the health care system [9].

Other authors such as Philippe Michel et al. showed that from a timeframe perspective, data collection methods can be grouped in two categories: cross sectional practices and the "time-series" practices. For the first method, the patient data was explored on a particular day, but only during that day. The second method involves a period of time for in-depth investigation. Data is collected at regular intervals by the doctor [10].

The latest data collection methods available in some European countries use an electronic real-time reporting system. This system is implemented over a wide area, in hospitals, care centres, laboratories, through the internet. Patients need to complete an online survey. The survey is accessible to every sick person that wants to participate. After the questionnaire is submitted, the data is directly associated with their medical records and stored into a secured database. Data can be accessed by the treating doctor [11].

In other regions of our planet, for example in the southern part of Africa, in 2013, the same electronic approach to patient safety measures was implemented on a large scale in Africa's developing economies: a **patient-event reporting system** called **PatSIS**. It was especially designed to manage injury-related events and so reduce the risk of error in healthcare centers. PatSIS represents a cross-competency solution to prevent and

monitor patient injury, to reduce time-consuming actions and paper-based processes [12]. A study conducted in 2005 in Korea showed that among the participating health care units, the large majority of them (more than 75%) were collecting medical mistake data manually, using paper, the verbal data collection method was applied in 30% of cases, and only less than 5% of them used an electronic reporting system [14]. The proposed solution to diminish the rate of occurrence of medical errors and unfavourable patient events and to increase the patient safety level is the large-scale implementation of IT-based techniques [13] [14].

Japan's central authorities introduced, in 2015, a new electronic medical data collection system that investigates and stores data about official deaths caused by medical procedures, the so-called "iryojiko chosa seido" (medical fatalities reporting system). All Japan's health care institutions (total of 180,000) were incorporated into the system and they were required to examine patient sudden deaths and report the results to the relevant independent institutions. The main shortcoming of this system is that it deals only with sudden patient deaths and not with the overall healthcare procedures. The hospital probation teams of a must demonstrate that a case represents an unexpected death. The independence of those teams is guaranteed by using external experts in the evaluation process [16]. In each Swedish state, for example, we find a different tumour documentation system; however, a national Cancer register, where every death is reported, also exists. A secondary sub-system of tumour data collection extracts elementary data about cancer tumours [17].

Germany's Fifth Book Social Code of 2014 established the introduction of an error notification system in the medical field. Medical units must have a risk management officer and a management system for patient complaints [18]. In Germany, there is a Critical Incident Reporting Systems (aka CIRS) which is organised in a well-developed network across the country [19]. This system allows the respondent (nurse, other health care worker, paramedic or doctor) to anonymously submit protected data via the internet, by completing a medical questionnaire about the medical critical event. He is allowed to add observations and recommendations to reduce the risk of recurrence of the event. The system focuses on medical events occurring in the medical care units, with the clear purpose of providing relevant multi-specialised schooling at a regional level and at an inter-professional degree. The CIRS Network has been extended internationally; it is currently being applied in Austria and Switzerland [20]. With respect to cancer health care units in Germany, for almost every type of cancer there are certified cancer centres which report their outcomes to a regional Cancer Registry (on a still non-mandatory basis) and to voluntary certification systems. Generally, there are many documentation systems which do not interconnect, generating several data comparability issues.

IV. CONCLUSIONS

The methods of data collection in hospitals depend on factors such as the final purpose of research, in terms of complexity and amount of data, and on the available instruments of collection, in terms of research staff, specific protocols, tools and costs of the overall process. These methods must extract relevant medical data in order to serve as an input for health care improvement programs, by granting users privacy, anonymity and data security.

Even though there are multiple options available worldwide when it comes to medical data collection methods, none of them represents the ideal solution. There are, in general, two types of data collection methods: manual methods and electronic/automated data collection systems. There are also traditional standard collection methods and recently we find more complex data collection tools. For example, surveys represent complex instruments of gathering statistically significant data, but they have a high cost of implementation.

Classic data collection approaches such as retrospective record reviews or record reviews of current inpatients, or staff interview of current inpatients, have shown their limitations in the modern world. Recently, electronic reporting system have started to be implemented in collecting patient data, medical errors and in the adverse events reporting phase. In order to efficiently reduce patient injury and the risk of medical errors, the reporting systems should provide free access to all relevant factors and they should allow anonymous completion of patient data or medical events. None of the quoted resources states anything about data comparability issues, from a time-series data perspective. When speaking about medical research or health care improvement plans, medical patient records are normally investigated on a year-on-year basis. Patient data is stored for several years for research and analysis purposes. Regardless of each specific method of data collection, the comparability of reported data must be always guaranteed, by every mean and every method used in the collection process.

The limitation of current research refers to unavailability of information regarding the following elements: patient storage data, transportation procedures for data between several locations, patient data privacy policy, and the anonymization techniques. There is no clear information about types of input data for different documents and forms, or about differences in terms of comparability between documentation systems. An analysis of similar data fields was not done. Other aspects refer to the input of patient data by medical staff into the documentation system, which inserts summary data without providing sufficient comprehensive details into the free text data fields. The current study does not treat the data input errors.

Future insights on this topic should address the standardization of different data collection methods in order to improve data comparability aspects.

REFERENCES

[1]   Minnesota Department of Health, Brief Overview of Data Collection Methods, St. Paul, MN 55164-0975, P.O. Box 64975, Minnesota, United States.

[2]   World Health Organization, WHO Patient Safety, Assessing and tackling patient harm: a methodological guide for data-poor hospitals, ISBN 978 92 4 150038 8, NLM classification: WX 173; Geneva, Switzerland, 2010.

[3]   Matthew B. Weinger, Jason Slagle, Sonia Jain, et al: Retrospective data collection and analytical techniques for patient safety studies, Journal of Biomedical Informatics, Volume 36, Issues 1–2, http://dx.doi.org/10.1016/j.jbi.2003.08.002, United States, 2003.

[4]   Ridyard, C. H. and Hughes, D. A., Methods for the Collection of Resource Use Data within Clinical Trials: A Systematic Review of Studies Funded by the UK Health Technology Assessment Program, Value Health, 13(8):867-872, doi: 10.1111/j.1524-4733.2010.00788.x, United Kingdom, 2010.

[5]   IOM (Institute of Medicine), Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement. Washington, DC: The National Academies Press, 2009, International Standard Book Number-13: 978-0-309-14012-6, United States.

[6]   Toronto Public Health, Mount Sinai Hospital, Centre for Addiction and Mental Health, St. Michael's Hospital: We ask because we care, The Tri-Hospital + TPH Health Equity Data Collection Research Project Report, St. Michael's Hospital Toronto, Ontario, Canada, 2013, http://www.stmichaelshospital.com/quality/data-collection-research-project.php.

[7]   J. Powell, Data sources and reporting, Birmingham & West Midlands Regional Cancer Registry, Queen Elizabeth Medical Centre, 29-42, International Agency for Research on Cancer (IARC), Birmingham B15 2TH, United Kingdom, https://www.iarc.fr/en/publications/pdfs-online/epi/sp95/sp95-chap5.pdf.

[8]   M.N. Sarkies, K.-A. Bowles, E.H. Skinner, et al: Data Collection Methods in Health Services Research. Hospital Length of Stay and Discharge Destination, ACI Applied Clinical Informatics, Vol. 6, No. 1, 2015, p. 96 – 109, doi:10.4338/ACI-2014-10-RA-0097.

[9]   American Hospital Association: Data Collection Methods, https://www.ahadataviewer.com/about/data/

[10]   Philippe Michel, Jean Luc Quenon, Anne Marie de Sarasqueta, et al: Comparison of three methods for estimating rates of adverse events and rates of preventable adverse events in acute care hospitals, BMJ 2004; 328:199, doi: https://doi.org/10.1136/bmj.328.7433.199, Pessac, France.

[11]   Trautmann F., Hentschel L., Hornemann B. et al.: Electronic real-time assessment of patient-reported outcomes in routine care – first findings and experiences from the implementation in a comprehensive cancer center, Supportive Care in Cancer, July 2016, Volume 24, Issue 7, pp 3047–3056, doi:10.1007/s00520-016-3127-0, Germany.

[12]   The Council for Health Service Accreditation of Southern Africa, PatSIS, 2013, http://www.cohsasa.co.za/patsis.

[13]   Thompson DA, Lubomski L, Holzmueller C, et al.: Integrating the intensive care unit safety reporting system with existing incident reporting systems, The Joint Commission Journal on Quality and Patient Safety, Volume 31, Issue 10, Pages 585–593, doi: http://dx.doi.org/10.1016/S1553-7250(05)31076-2, 2005.

[14]   Jeongeun Kim, David W. Bates: Results of a survey on medical error reporting systems in Korean hospitals, International Journal of Medical Informatics, Volume 75, Issue 2, February 2006, Pages 148-155, doi: http://dx.doi.org/10.1016/j.ijmedinf.2005.06.005.

[15]   Canadian Association of Provincial Cancer Agencies, Incident Reporting Systems in Canada, https://www.cihi.ca/en/national-system-for-incident-reporting-home-page.

[16]   Tomoko Otake: Shortcomings identified in new reporting system into hospital deaths, Japan, http://www.japantimes.co.jp/news/2015/10/05/reference/shortcomings-identified-new-reporting-system-hospital-deaths/.

[17]   Mieke Van Hemelrijck, Annette Wigertz, Fredrik Sandin, et al: Cohort Profile: The National Prostate Cancer Register of Sweden and Prostate Cancer data, Base Sweden 2.0, Int J Epidemiol. 2013 Aug;42(4):956-67, doi:10.1093/ije/dys068, Sweden.

[18]   Jürgen Brenn, Fehlermeldesystem CIRS-NRW auf einem guten Weg, In: Rheinisches Ärzteblatt. Ärztekammer Nordrhein und Kassenärztliche Vereinigung Nordrhein. S. 21-22, January 2014, Germany.

[19]   Krankenhaus-CIRS-Netz Germany, http://www.kh-cirs.de/.

[20]   Wikipedia, Critical Incident Reporting System, https://de.wikipedia.org/wiki/Critical_Incident_Reporting_System, 2017.