A narrative systematic review of medical registries.

Williams ZR, Karpelowsky J*

The Children's Hospital, University of Sydney, Westmead, Australia

Abstract

Introduction: Medical registries are valuable stores of data for research and public health monitoring purposes. This review aims to identify the benefits of medical registries to healthcare, determine the components of effective registries, assess concerns with patient privacy, and develop guidelines for designing quality medical registries.

Method: Pubmed was screened for articles related to medical registries, data management, and patient privacy. Thirty-seven articles adhered to the selection criteria and were included in this review.

Results: Benefits of medical registries: Registries facilitate health outcomes research and can improve patient prognosis by identifying effective care protocols. Data quality and error reduction: Data quality control rests upon clear and standardized processes for data collection and input. Error reduction can be achieved by reducing manual text input and automating data transfer between sources. Automatic feedback features in registries are the most feasible methods of identifying data error. Privacy: Patient privacy must be preserved by anonymizing or pseudonymizing registry data. The preservation of public trust is important to maintain participation in opt-out registries.

Discussion and Conclusion: Registries can be designed to promote high quality data collection and storage by reducing systematic and random error. This review presents a framework for designing effective medical registries that fulfil research and health monitoring purposes while respecting patient privacy.

Keywords: Data quality, Error reduction, Cystic fibrosis, Myeloma, Cancer registries.

Accepted on February 08, 2019

Introduction

Medical registries are databases of patient-specific health information [1-3]. A quality registry is a valuable tool for collecting, storing, and processing patient data. Registry data are analysed for set purposes including epidemiological, aetiological and health outcomes research, analysis of healthcare protocol inconsistencies across institutions, and assessment of treatment efficacy. The data collected in registries aid patients, practitioners, and policy makers by providing regular feedback on areas in which healthcare can be improved [1-9].

Quality data in registries are necessary for effective analysis [2,7,8,10]. Errors in data collection, storage and processing must be regularly monitored. Strategies for error reduction and data quality maintenance should be implemented into registry design [2,7].

Medical registries raise concerns regarding data mishandling and patient privacy [11-18]. Registries must balance quality of data with the privacy of patients, without compromising the ability for research to be conducted [14,17,18].

This review will outline the components of quality medical

registries and the benefits these registries have on patients, medical practitioners and policy makers. Guidelines will then be provided from which such a registry can be developed and utilized whilst maintaining patient privacy.

Literature Review

A systematic search of Pubmed was performed using combinations of "medical registry", "health", "outcomes", "data quality", "data management", "design", "ehealth records", and "privacy". Reference lists of articles highlighted by this search were consulted and relevant articles were considered.

Inclusion criteria consisted of: Articles published in English; and articles discussing health-related registries, registry design, data management concerns, methods to reduce data error, methods to improve registry data quality, or strategies to preserve patient privacy in medical record keeping. Thirty-seven articles adhered to these criteria.

Articles were reviewed under key themes: Benefits of medical registries to healthcare, improving data quality, reducing data error, implementing data quality feedback mechanisms, and patient privacy concerns.

Results

Healthcare benefits of medical registries

Data stored in medical registries provide useful information to practitioners and health policy makers. Analysis of registry data is used for research and quality control purposes to improve patient care and treatment outcomes.

Health outcomes research

Registries provide a database from which key patient outcomes for conditions or procedures can be easily compared [9]. The consistency of the data collected in a registry allows key outcome measures to be compared efficiently between patients, practitioners, and institutions. The American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) was established to collect data on risk-adjusted surgical outcomes and compare the outcomes at the Department of Veteran's Affairs hospitals with the national average of the United States. The development of an accurate database allowed for this comparison between the quality of surgical care in civilian and veteran hospitals [19-22]. Medical registries hence serve as storage sites for patient data and facilitate comparisons of health outcomes [5-8,19].

Variations in outcomes

Registries enable enquiries into the determinants of variances in patient outcomes by identifying outliers in the data [20]. A study using ACS-NSQIP databases compared outlier status between 2005 and 2007. Of the analysed institutions, 89% improved on morbidity outcomes and 80% on mortality outcomes [20]. The rapid feedback facilitated by registries provides instruction regarding which institutions require further support to improve in outcome measures [21].

Variations in care practices

Registries often include data on both care practices and outcomes. This allows for associations between specific care practices and positive health outcomes [4-6]. Promising care practices can be identified and refined. This may drive standardization of diagnostic and therapeutic practices across institutions. The efficacy of these protocol changes can be reassessed after secondary comparison of protocols and outcomes with international standards to implement optimal care protocols [1,2,6-8]. Care processes at institutions with improved patient outcomes in the ACS-NSQIP database share their processes with a central body, which disseminates the information to other institutions. Reviews of practices at under-performing institutions are also conducted to identify areas with potential for improvement [19,22]. As a result of NSQIP data feedback, Veteran's Affairs hospitals reduced 30-day postoperative mortality by 47% and 30-day postoperative morbidity by 43% from 1991-2008 [22]. The reliable data stored in registries facilitate assessments of care practices between practitioners and institutions, which improve patient outcomes [4-6].

International comparisons

Registries facilitate comparisons of disease incidence, efficacy of preventative measures, and therapeutic outcomes [9]. Registries collecting similar data sets in two countries can be studied to compare healthcare protocols and patient outcomes internationally. A study comparing the US Cystic Fibrosis Foundation Patient Registry and the Australasian CF Data Registry concluded that the positive effect of newborn cystic fibrosis diagnosis by screening rather than clinical diagnosis on pediatric outcomes, specifically lung function and BMI, was significantly less in Australia than in the United States [23]. Patients in Australia were prescribed pancreatic enzymes less frequently than patients in the United States, highlighting a potential area of under treatment in Australia. Australia was reported to treat pulmonary exacerbations more aggressively than the United States, an approach associated with improved clinical outcomes [23]. International comparisons of registries promote more extensive research into the efficacy of screening and treatment protocols, the results of which will be applied to improve patient outcomes globally [9].

Self-evaluation of practitioners and institutions

Private evaluation of personal performance against benchmarks is often possible in medical registries [8]. In the NSQIP database, institutions are assigned a unique code to allow private comparison of their risk-adjusted surgical outcomes to national averages [19]. The feedback provided by registry data analysis allows for identification of areas of improvement and encourages collaboration. This drives refinement of institutional practices which improves patient outcomes, while ensuring varying levels of performance do not become public knowledge [8,21,22].

Healthcare organisation

Registries allow monitoring of changing health problems [1]. Registry data can inform health policy by identifying areas that increasingly burden the healthcare system. Prevention strategies and therapeutic changes can then be implemented to address these areas [1]. Registry data can be used to quantify disease prevalence and treatment costs to inform government and institutional resource allocation [1]. Registry data can facilitate analysis of the efficacy of therapeutic or publichealth interventions to inform government decisions regarding the subsidisation of pharmaceuticals or health services [9]. Outcome measures identified from registry data can also highlight research areas with the greatest potential and inform public health spending [1,9].

Strengthening of statistical analysis

The incorporation of data from multiple centres in one registry allows for stronger statistical analysis than studies of single institutions. Centralizing data has particular benefits for registries of rare diseases, of which an institution may see only a few cases per year [14]. Without registries, data from small centres are often not included in research. Centralizing data in multi-institutional registries overcomes this bias in data collection [14]. Larger data sources facilitate effective research and provide more accurate overviews of health problems [4,5,14]. Population-based registries provide a breadth of data that improves the validity of studies. These registries significantly reduce the cost of population-based studies as data has already been collected and processed [24].

Data Quality

Registries require quality data to function effectively and

produce a reliable output [6-8,10]. The quality of data relates to the ability of the data to contribute the purpose of the registry [7]. Templates have been developed to quantify data quality based on intrinsic, contextual and representational attributes. These metrics are independent of context and provide a method of standardizing data quality analysis across registries [8,10,25]. Measuring the quality of registry data allows for assessment of the effectiveness of a registry and identification of areas in which data quality could be improved.

Intrinsic data quality relates to completeness, accuracy and consistency [10]. Data completeness describes the absence of blank fields for which data are available [2,7]. Accuracy describes the degree to which data are reliable, objective and correct [8,10]. Consistency describes whether data correspond between registries and the original source, that is the data has not changed as a result of transcription error [8].

Contextual data quality describes the relevance and timeliness of data. Relevance relates to the usefulness of data for the purpose of the registry. Timeliness describes whether the data are sufficiently current to serve this purpose [10]. Some registries collate data for assessment before and after alterations in care protocols. Such registries would require data over a particular time period. Timeliness describes the appropriateness of the range of dates to which the data corresponds [10].

Representational data quality refers to the accessibility and comprehensibility of data [10]. Granularity is a marker of representational quality and describes the appropriateness of the level of detail. Data that are too detailed are irrelevant and can be incomprehensible. Data of insufficient detail do not serve the purpose of the registry [10].

Data Errors

Quality data are absent of error [7]. Data errors significantly alter the output of registry data analysis, particularly for registries of rare diseases with fewer data points. Reducing error is vital to maintaining the integrity and reliability of the registry [7].

Systematic data errors refer to errors in registry design. These errors include programming mistakes, the collection of data that do not serve the purpose of the registry, and vague data collection instructions or registry field descriptions that result in unstandardized responses [7]. Random data errors refer to errors in data transference from the source to the registry. These errors include inputting data into the wrong field, typing errors, and the mixing of data between cases [7]. Data can also be inaccurately recorded at the level of patient records, resulting in subsequent errors in registries [2,26,27]. Most data errors can be attributed to programming errors [2]. However, transcription errors, missing data, unsuitable granularity, and changes in thresholds or definitions of disease benchmarks over time also represent considerable sources of registry data error [10].

Controlling Data Quality

Data quality control rests upon the existence of clear and standardised protocols for data collection and data input [2,7].

Data collection

Data should be collected close to the original source. Collection should be conducted as soon as data are available to allow

clarification if required [2]. Data collectors should be trained centrally in collection protocols to standardise collection practices across institutions [2]. Training should involve clarification of the purpose of the registry and development of database literacy [7]. The success of the NSQIP database is in part attributed to a fulltime data collector and reviewer who are trained in NSQIP collection and input methods and works independently to participating surgeons [20]. Data collection control is integral to maintaining data quality.

Registry design

The database should be user-friendly and require minimal training to navigate [7]. Data fields should provide clear definitions with links to further clarification if required. Reducing ambiguity in data fields is imperative for improving data quality [2,7,10]. Data fields should only be included if they contribute to the purpose of the registry and are objectively measurable [10]. The registry should not contain text inputs. Alternatives include check boxes and drop-down menus for categorical variables. For numerical fields, a slider function should be used where appropriate to reduce transcription errors [2].

Reducing transcription error

Manual transcription and transference of data between sources increases the likelihood of data error. Error is reported to be as high as 27% when data is entered in duplicate [7,28]. Establishing automatic transfer of data from electronic health records to registry fields reduces transcription error. Record linkage also avoids data duplication by connecting patient data from multiple registries under a common identifier [5,10,29,30]. Record linkage was used effectively in the Netherlands to link breast cancer screening and cancer registries [29]. The cancer registry allowed for comparison of breast cancer incidence before and after implementation of the screening program. Record linkage reduced error by automating transcription and flagging data duplication [29].

Data Quality Feedback

Feedback on data quality can be used to assess data quality standards and identify sources of data error [2]. The data collection and input processes can then be streamlined to reduce error and improve the quality of data [6].

Data audits are only effective if conducted regularly during the data input process [2,7]. Whilst onsite data verification throughout the input process is useful, it is costly and unfeasible. Repeating the data entry process with another staff member and assessing transcription error is also costly and impractical [2,10,31]. Visual checks of data summary pages often overlook inconspicuous errors which may significantly affect analysis [2].

Automatic feedback features built into a database are more effective and feasible alternatives to data audits [10]. Reports can be generated to detect outliers, identify blank fields and flag areas of potential duplication [10]. Automatic identification of anomalies can be achieved by implementing strict ranges for registry fields [2]. This is only effective, however, for fields with small ranges. Automatic comparison of the spread of the data with similar data sets in independent registries would also be beneficial [2]. Record linkage between electronic health records Table 1. Suggestions for registry design and implementation.

Data Quality: Registry Design Features
All registry fields should contribute to the purpose of the registry. No extraneous information should be collected.
The number of fields should be kept to a minimum.
Data field descriptions should be concise and unambiguous.
Drop down menus and check boxes should be used in place of text input.
Slider functions should be used in place of numerical inputs.
Strict ranges should be set for slider functions or numerical inputs.
Fields should be labelled as mandatory or optional to allow data entry to the registry when less important details are not available. Mandatory fields should be those that are necessary to fulfil the registry's primary purpose.
Automatic feedback features such as regular report generation on outliers, the number of blank fields, and data duplication should be incorporated into the registry design.
Data dictionaries should be standardised for basic information to allow automatic and accurate transfer of data from electronic health records and between registries.
Registries for individual institutions should be linked with electronic health records where possible to reduce data transcription error.
Data Quality: Data Collection and Input
Registry database should be user friendly.
Staff should be trained centrally to standardise collection and input protocols where automatic feedback is not feasible.
Patient Privacy Protection
Data should be anonymised or pseudonymised before research is conducted.
Unencrypted data should only be accessed or adjusted by primary data collectors.
An opt-out participation process should be implemented.

and across registries would also ensure consistency across files and allow quicker data transcription. Record linkage, however, has implications for individual privacy as data fluidity between registries can compromise security [32].

Privacy

There is vocal concern at a government and public level over the protection of individual privacy and disclosure of sensitive data contained in medical registries [11-18]. Collected data must only be used to fulfil the purpose of the registry [11]. Access to this data must be restricted to individuals involved in data collection or analysis [10]. Whilst measures can be taken to dissociate patient identifiers from their personal data, these measures must not alter or encode the data to a point where the registry can no longer serve its intended purpose. Trade-offs are made between privacy protection and usefulness of data for research to ensure a valuable output whilst maintaining public trust [10].

Anonymization and pseudonymization are practical forms of privacy protection [10,13]. Anonymization describes the encryption of identifiable data points. Anonymized data cannot be linked back to an individual. K-anonymization is a useful tool in smaller registries for rare medical conditions. In these registries, individuals are identifiable by many data points given the small number of participants [10]. K-anonymization encrypts enough data points so as to deidentify individuals completely from their data. However, k-anonymity can be too restrictive and can compromise data analysis in some registries [12,33]. Pseudonymization allows data to be traced back to individuals after multiple steps of decoding [10]. This is particularly useful for registries that continue to collect patient data during follow up consultations [15].

Access to data should be restricted to the data collector and protected by certificate-based authentication [10]. Software should then be used for encryption before research is commenced. Any corrections made to the data should be conducted by the principal data collector to protect the identity of the patient from the researcher [10].

The enforcement of privacy protection policies is important in medical registries to maintain public trust and promote high rates of participation. Patient consent is usually required to include patient data in registries [34]. However, actively seeking consent results in low levels of participation [14,17]. Opt-out policies are frequently approved by ethics committees if the benefits of the registry to the public are determined to outweigh the impingement on personal privacy [35]. Opt-out consent policies result in higher registry participation [17]. In the opt-out Victorian State Trauma Registry, only 0.5% eligible patients declined to participate [17]. The inclusion of all patient data is imperative to ensure unbiased data and to produce quality research [14,34]. Incomplete registries, particularly when monitoring rare diseases or rare responses to treatment, can cause researchers to overlook patterns of side effects or causal links that occur in a small proportion of the population [14]. Full participation also allows data to be adjusted for risk factors before being used for further analysis [17]. High participation is thus necessary to ensure optimal usefulness of registry research, but must be balanced with consideration of patient autonomy [14].

Registry data must only be used for the registry's intended purpose [11]. A primary public concern is the release of patient data for marketing, insurance, or commercial purposes [18]. When the purpose of a registry is to monitor the efficacy of a commercial device, patients should be informed of this purpose and given the opportunity to decline participation. At times, providing data or study results to a commercial manufacturer may be necessary to improve patient outcomes, but patient information must remain protected. Researchers using registry data must be transparent about their purposes, and any intent for commercial partnership [18]. Public trust must be maintained with transparency as loss of public trust will lower participation and be of detriment to scientific advances [36]. Privacy protection is thus imperative and must be navigated without weakening scientific research.

Discussion

This review presented research into the makeup of effective medical registries. Registries can be designed to control intrinsic, contextual, and representational data quality and reduce data error (Table 1). Database design and data collection protocols must be continually developed to improve data quality and maintain the integrity of the registry.

A balance between patient privacy and bulk data collection for research and healthcare purposes must be respected when establishing medical registries [10]. The collection of extraneous information along with key patient data points is a breach of patient privacy and compromises the ethical integrity of the registry. Maintaining public and government trust in the security of patient data is imperative if registries are to continue to be used for research and healthcare monitoring purposes. Collected data must be secure, encrypted and analysed only for the purpose of the medical registry [11].

Automatic data transfer from electronic health records to medical registries is an effective method of reducing transcription error and avoiding data duplication. Unique national medical identifiers can link hospital and non-hospital data and negate the repetition of data collection and input [2,5,13,16]. Medical record numbers in Australian institutions can be used to integrate data from electronic health records into institution-specific registries. However, in the absence of a system of national health identification numbers, fluid data transfer between registries or into centralised, multi-institutional databases remains unfeasible. The Australian and New Zealand (ANZ) Myeloma and Related Diseases Registry monitors participation and by linkage with state and national cancer registries. Cases missing from the ANZ Myeloma and Related Diseases Registry are identified to encourage close to 100% participation and ensure population-wide data is included in the registry [35]. However, comparison of patient participation is far from the potential of fluid record linkage between registries. Complete data transfer would save time and reduce costs in data collection [37].

Conclusion

Data security can be compromised with fluid data transfer between registries and patient records as a greater number of staff have access to identifiable data. While the integration of electronic health records improves data quality, patient privacy is difficult to ensure. A balance must be achieved between measures taken to respect the privacy of patient data and measures which improve the quality of data and the efficacy of the registry.

Medical registries provide an effective storage site for patient data for use in research and healthcare evaluation. By reviewing the components contributing to the effectiveness of medical registries, this review provides guidelines for developing quality databases from which effective research and analysis can be performed.

References

- Oberbichler S, Hrbst A. Functional requirements regarding medical registries: Preliminary results. InICIMTH. 2013;190:112-4.
- 2. Arts DG, De-Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: A literature review, case study, and generic framework. J Am Med Inform Assoc. 2002;9(6):600-11.
- 3. Drolet BC, Johnson KB. Categorizing the world of registries.

J Biomed Inform. 2008;41(6):1009-20.

- Plotnicki L, Kohl CD, Hocker B, et al. The certain registry: A novel, web-based registry and research platform for pediatric renal transplantation in Europe. Transplant Proc. 2013;45(4):1414-7.
- 5. Vener DF, Gaies M, Jacobs JP, et al. Clinical databases and registries in congenital and pediatric cardiac surgery, cardiology, critical care, and anesthesiology worldwide. World J Pediatr Congenit Heart Surg. 2017;8(1):77-87.
- 6. Veer SN, Keizer NF, Ravelli AC, et al. Improving quality of care: A systematic review on how medical registries provide information feedback to health care providers. Int J Med Inform. 2010;79(5):305-23.
- 7. Pass HI. Medical registries: Continued attempts for robust quality data. J Thorac Oncol. 2010;5:S198-9.
- 8. Salati M, Brunelli A, Dahan M, et al. Task-independent metrics to assess the data quality of medical registries using the European Society of Thoracic Surgeons (ESTS) database. Eur J Cardiothorac Surg. 2011;40(1):91-8.
- 9. Jefford M, Stockler MR, Tattersall MH. Outcomes research: What is it and why does it matter?. Internal Med J. 2003;33(3):110-8.
- 10. Sariyar M, Borg A, Heidinger O, et al. A practical framework for data management processes and their evaluation in population-based medical registries. Inform Health Soc Care. 2013;38(2):104-19.
- 11. Ameron P, Finch C, Gabbe B, et al. Developing Australia's first statewide trauma registry: What are the lessons?. ANZ J Surg. 2004;74:424-8.
- 12. El-Emam K, Dankar FK. Protecting privacy using k-anonymity. J Am Med Inform Assoc. 2008;15(5):627-37.
- Gardner J, Xiong L. An integrated framework for deidentifying unstructured medical data. Data Knowled Eng. 2009;68(12):1441-51.
- Hansson MG, Simonsson B, Feltelius N, et al. Medical registries represent vital patient interests and should not be dismantled by stricter regulation. Cancer Epidemiol. 2012;36(6):575-8.
- 15. Neubauer T, Heurix J. A methodology for the pseudonymization of medical data. Int J Med Inform. 2011;80(3):190-204.
- 16. Sandea IA. An analysis of the legal aspects concerning data protection in electronic medical registry. Applied Medical Informatics. 2009;25(3-4):16-20.
- 17. Williamson OD, Cameron PA, Neil JJ. Medical registry governance and patient privacy. Med J Aust. 2004;181(3):125-6.
- 18. Hansson M. Where should we draw the line between quality of care and other ethical concerns related to medical registries and biobanks?. Theor Med Bioeth. 2012;33(4):313-23.
- 19. Khuri SF, Daley J, Henderson W, et al. The department of veterans affairs' NSQIP: The first national, validated,

5

outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. Ann Surg. 1998;228(4):491-507.

- 20. Hall BL, Hamilton BH, Richards K, et al. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: An evaluation of all participating hospitals. Ann Surg. 2009;250(3):363-76.
- 21. Khuri SF. The NSQIP: A new frontier in surgery. Surgery. 2005;138(5):837-43.
- 22. Khuri SF, Henderson WG, Daley J, et al. Successful implementation of the department of veterans affairs' national surgical quality improvement program in the private sector: The patient safety in surgery study. Ann Surg. 2008;248(2):329-36.
- 23. Martin B, Schechter MS, Jaffe A, et al. Comparison of the US and Australian cystic fibrosis registries: The impact of newborn screening. Pediatrics. 2012;129(2):e348-e55.
- 24. Thong MS, Mols F, Stein KD, et al. Population-based cancer registries for quality-of-life research: A work-in-progress resource for survivorship studies?. Cancer. 2013;119(S11):2109-23.
- 25. Pipino LL, Lee YW, Wang RY. Data quality assessment. Communications of the ACM. 2002;45(4).
- 26. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. 1997;4(5):342-55.
- 27. Aronsky D, Haug PJ. Assessing the quality of clinical data in a computer-based record for calculating the pneumonia severity index. J Am Med Inform Assoc. 2000;7(1):55-65.
- 28. Goldberg SI, Niemierko A, Turchin A. Analysis of data

errors in clinical research databases. AMIA Annu Symp Proc. 2008;pp:242-6.

- 29. Schouten LJ, Rijke JM, Schlangen JT, et al. Evaluation of the effect of breast cancer screening by record linkage with the cancer registry, The Netherlands. J Med Screen. 1998;5(1):37-41.
- Christen P, Goiser K. Quality and complexity measures for data linkage and deduplication. Studies in Computational Intelligence. 2007;pp:127-51.
- Willis CD, Jolley DJ, Neil JJ, et al. Identifying and improving unreliable items in registries through data auditing. Int J Qual Health Care. 2011;23(3):317-23.
- 32. Kluding PM, Denton J, Jamison TR, et al. Frontiers: Integration of a research participant registry with medical clinic registration and electronic health records. Clin Transl Sci. 2015;8(5):405-11.
- Lei J. Information and communication technology in health care-do we need feedback?. International Journal of Medical Informatics. 2002;pp:75-83.
- 34. Tu JV, Willison DJ, Silver FL, et al. Impracticability of informed consent in the registry of the Canadian stroke network. N Engl J Med. 2004;350(14):1414-21.
- 35. Bergin K, Moore E, Quilten Z, et al. Design and development of the Australian and New Zealand (ANZ) myeloma and related diseases registry. BMC Med Res Methodol. 2016;16(1):151.
- 36. Myhre SL, Kaye J, Bygrave LA, et al. eRegistries: Governance for electronic maternal and child health registries. BMC Pregnancy Childbirth. 2016;16(1):279.
- Karen G. Registries and paediatric radiation oncology. J Med Imaging Radiat Oncol. 2014;58(6):691-2.

*Correspondence to:

Zoe Rose Williams The Children's Hospital University of Sydney Westmead, Australia Tel: +61411951001 E-mail: zwil7748@uni.sydney.edu.au; zrwilliams321@gmail.com