



Published in final edited form as:

Appl Nurs Res. 2012 May ; 25(2): 108–116. doi:10.1016/j.apnr.2010.02.004.

Research strategies that result in optimal data collection from the patient medical record

Katherine E. Gregory, PhD, RN^{a,*} and Lucy Radovinsky, MSN, RN^b

^a W.F. Connell School of Nursing, Boston College, Chestnut Hill, MA 02467, USA

^b Joseph M. Smith Community Health Center

Abstract

Data obtained from the patient medical record are often a component of clinical research led by nurse investigators. The rigor of the data collection methods correlates to the reliability of the data and, ultimately, the analytical outcome of the study. Research strategies for reliable data collection from the patient medical record include the development of a precise data collection tool, the use of a coding manual, and ongoing communication with research staff.

1. Introduction

The patient medical record is a rich source of data frequently used by nurses conducting clinical research (Gearing, Mian, Barber, & Ickowicz, 2006; Gregory, 2007; Halm et al., 2009; Johnson, Krosnick, Carson, McDade, & Laraway, 1998; Jones & Fennie, 2007). The patient medical record is often used as a primary source of retrospective data for the purposes of epidemiological analysis and is considered to be the gold standard in any study to identify demographic factors, clinical data variables, specific aspects related to treatment regimens, and ultimately patient mortality and morbidity (Cassidy, Marsh, Holleran, & Ruhl, 2002; Murray et al., 2003). Data collection from the medical record involves reviewing specific sources within the record. These include nursing, physician, and consultation notes; admission and discharge reports; laboratory and diagnostic test reports; surgical reports; and other clinical and administrative documentation. This is not easily done. A sound strategy prior to initiating the data collection effort is required. This article provides a description of strategies that may be implemented to ensure optimal data collection from the medical record. These research strategies have been tested and successfully demonstrated in both retrospective and prospective data collection efforts (Eder, Fullerton, Benroth, & Lindsay, 2005; Gearing et al., 2006; Pan, Fergusson, Schweitzer & Hebert, 2005). The proposed strategies will provide nurse investigators with specific methods that will contribute to the successful collection of data from the patient medical record by research staff using a data collection tool.

2. The utility of patient medical record data in clinical research

Patient medical record data collected via retrospective chart review contribute to at least 25% of the scientific articles published in clinical journals (Gilbert, Lowenstein, Koziol-McLain, Barta, & Steinder, 1996; Murray et al., 2003; Worster & Haines, 2004). Retrospective medical record review is a useful methodology in exploring research questions that often cannot be answered using prospective trials. These questions may

*Corresponding author. Brigham and Women's Hospital, Boston, MA 02115, USA. Tel.: +1 617 552 4886(Office), +1 781 777 1257(Home); fax: +1 617 552 0745. katherine.gregory.2@bc.edu (K.E. Gregory).

include the study of potentially harmful effects of exposures, as well as the occurrence of rare events in exposures when randomization is not possible (Worster & Haines, 2004). Due to ethical implications associated with research involving vulnerable participants, medical record review is a useful method for exploring questions in perinatal research (Battle, Zlotnick, Miller, Pearlstein, & Howard, 2006), neonatal research (Fanaroff, Wilson-Costello, Newman, Montpetite, & Fanaroff, 2006; Sharek et al., 2006), pediatric research (Gunn, Hansen, & Kaelber, 2007; Taylor et al., 2006), and mental health research (Biederman et al., 2005; Staller, 2004). The use of medical record data is highly practical in the study of longitudinal disease pathogenesis over an extensive period (Bates et al., 2006; Gaudreau, Gagnon, Roy, Harel, & Tremblay, 2007), in the study of quality assurance and improvement in clinical practice (Colón-Emeric et al., 2006; Scanlon, Miller, Harris, Schulz, & Sedman, 2006), and in pilot studies providing data for future clinical research initiatives (DeMille, Deming, Lupinacci, & Jacobs, 2006; Haley et al., 1980; Pronovost et al., 2003). Finally, data collected from the medical record are essential to establishing a health history on patients who provide biologic sample for the purposes of cutting-edge disease prediction research. Clinical nurse investigators working across varied patient populations aiming to predict disease, prevent disease, or promote health are likely to ask questions that require data from the medical record.

The advantages of using data obtained from the medical record via retrospective record review include the ability to access large amounts of clinical data at a relatively low cost, the ability to study associations between exposure and disease over long periods, and the ability to evaluate hypotheses pertaining to clinical research questions, which may then be tested using prospective trials. Limitations of using data obtained from the medical record via retrospective review include incomplete or missing data within the medical record, records lacking specific patient information, difficulty in interpreting or verifying documented information, and variability in the quality of documentation among health care personnel (Gearing et al., 2006). A sound methodology and a stepwise approach to collecting data from the medical record are important in the conduct of clinical nursing research. The strategies presented will capitalize on the advantages of using medical record data and minimize the limitations of using this source of data to explore issues relevant to clinical nursing research.

3. Collecting patient medical record data for research

Protocols and guidelines for data abstraction from the medical record must be developed prior to launching any medical record data collection effort. Strategies underpinning these protocols and guidelines include (a) the development and testing of a data collection tool; (b) the use of a coding manual guiding data collection of specific variables from the medical record; and (c) the selection, training, management, and ongoing communication with research staff (i.e., data abstractors). The research strategies presented pertain to these three essential elements for data abstraction from the medical record (Eder et al., 2005; Gearing et al., 2006; Jansen et al., 2005; Pan et al., 2005).

3.1. The development and testing of a data collection tool

The development of a data collection tool begins with the investigator's assessment of whether the necessary data are available in the patient medical record. This assessment is most often based on previous experience and in collaboration with clinical staff, ideally nurses, who document most frequently in the medical record (Eder et al., 2005). When developing the tool to effectively and efficiently collect data from the medical record, the investigator should consider the organization and structure of the data collection tool, the detailed nature of the data to be collected from the medical record, and finally, whether a

paper-based or electronic version of the data collection tool should be incorporated into the study methodology (Gearing et al., 2006).

3.1.1. Organization and structure of the data collection tool—Logical organization and a structure that is user-friendly are simple but important strategies for ensuring optimal data collection from the medical record (Gearing et al., 2006). A conceptual framework underpinning the study that includes the variables of interest is often helpful in defining the organizational structure and framework of the data collection tool. The organization and structure of the data collection tool are important to the actual data collection because ideally, the tool will correspond to the data as it appears in the patient record. For example, demographic data should be grouped together, followed by specific clinical data that are reported collectively in specific sections of the medical record (i.e., study variables pertaining to laboratory values and pathological reports, radiology reports, and surgical reports). When collecting longitudinal data that are reported on a daily patient flow sheet or in the physician or nursing notes, investigators should design the data collection tool in a systematic manner by date and group together by physiological system. Organization of the data collection tool in this manner facilitates reliable data collection and ease of use by the data abstractor. The ease and efficiency of data collection from the medical record will correlate with the degree to which the tool corresponds to the data as these appear in the medical record.

The data collection tool must be systematically organized and easy to use. The tool may employ the use of charts and tables to organize the data collection. Use of the data collection tool is further enhanced by easy-to-read fonts such as 12-point Ariel or 10-point Verdana and adequate spacing for documenting data collected from the medical record (Bernard, Lida, Riley, Hackler, & Janzen, 2002). As in the development of any document, readability is important (Coyne et al., 2003). Formatting that includes the adequate use of white space and right justification of documentation areas on the data collection form not only is more efficient at the time of data collection but also facilitates more accurate and efficient data entry from paper-based tools (Coyne et al., 2003; Gearing et al., 2006). Finally, investigators must ensure that the data collection tool provides a designated space for documentation of the study ID number on each page and the name or initials of the data abstractor and the date that the data were abstracted from the medical record. These are critical factors, especially when investigators are using a paper-based tool greater than one page, as pages may become separated or lost. These are all effective strategies not only at the time of data collection but also at the time of data entry.

Pilot testing a data collection tool early in the design phase of the study will help investigators gauge the ease of use and efficiency of the data collection. Due to specific insights on how information is documented in the medical record, clinical nurses are the ideal consultants to team with investigators at this stage of developing and testing a data collection tool. Fig. 1 provides an exemplary template of a data collection tool. This data collection tool, and the corresponding code book, seen in Fig. 2, has been developed for illustrative purposes only.

3.1.2. Detailed nature of the data to be collected from the medical record—To ensure validity and accuracy of data collection, variables should be recorded on the data collection tool with a simple, unequivocal response. When numerical data are not available or appropriate for a specific study variable, a categorical response may be assigned. These types of responses may be used to capture surgical or radiological findings, as well as for missing or undetermined variables. The categories must be clearly defined on the data collection tool for the research staff collecting data from the medical record. These strategies aimed at standardizing the detailed nature of the data collected are important to the internal

validity and reproducibility of any retrospective study (Jansen et al., 2005). All of these strategies pertaining to the detailed nature of the data to be collected from the medical record ensure the validity of the data and, in turn, rigor of the study methodology and ultimately, the findings reported (Allison et al., 2000).

Investigators with limited experience or seeking to expedite the data collection effort may be tempted to simplify data collection and subsequent statistical analysis by collecting data from the medical record as categorical or dichotomous variables. A key strategy to research success using medical record data is to collect the most precise and detailed data possible from the medical record. Variables can be collapsed and consolidated into categorical or dichotomous variables at a later date, prior to analysis as necessary. If the investigator collects data from the medical record that are not representative of the original precise data available on each variable and later want to analyze data at that level of specificity, he or she will be forced to recollect data from each medical record for each participant. This is not only inefficient but can also be a costly and time-consuming endeavor that is preventable by collecting detailed data at the initial time of data collection.

3.1.3. Paper or electronic data collection methods—The data collection tool may be a paper document or an electronic record (Allison et al., 2000). Both types of data collection tools have attributes and limitations. A paper document is often more cost-effective and easier to utilize at the point of data collection, typically in the medical records department across multiple sites (Allison et al., 2000). However, the use of a paper document requires that data then be entered into an electronic database for analysis. Even under the best of circumstances, the data entry process is fraught with opportunity for error, which will influence analyses and study findings. Electronic data collection directly into a database decreases the opportunity for error in data entry, which can result in more reliable data collection and subsequent analysis (Worster & Haines, 2004). In addition, electronic data collection facilitates centralization and access to data and can be more cost-effective in large investigations (Gearing et al., 2006).

Data collection via electronic means requires that each research team member has access to an electronic device to collect and enter data. This may be feasible in some settings where networked computers are available to medical record abstractors and a shared database can be accessed and kept secure. It may also be feasible when a study has adequate resources to provide electronic devices such as a laptop computer or a personal data assistant to each research team member collecting data from the medical record. Data collection from the medical record via electronic means may be the gold standard of the future; however, it is yet to be the norm for most studies to date (Burt, Hing, & Woodwell, 2006; Faster Cures: The Center for Accelerated Medical Solutions, 2004; Murphy, Ferris, & O'Donnell, 2008).

3.2. The use of a coding manual guiding data collection from the medical record

The development and use of a coding manual indicating the best source of information within the medical record for each study variable is an essential strategy for data collection from the medical record. The coding manual provides clear direction to the individuals abstracting data from the medical record on the protocol pertaining to the data collection for each study variable included on the data collection tool. It lists each variable with an explanation as to how the variable should be recorded on the data collection tool (Gearing et al., 2006). In addition, the coding manual provides a description of the best source of information within the medical record for each variable included in the study and any other relevant information necessary to identify and record the data. In short, data abstractors use a coding manual as a guide in their work. Fig. 2 provides an example of a coding manual, which has been developed for illustrative purposes only.

A coding manual may be used to direct the data abstractor to the paper medical record for collection of some of the study variables and the electronic medical record for collection of other study variables. This becomes important as the best source of data for each of the study variables often varies depending on the year that the patient was cared for, the institution where the patient was cared for, and even the admitting clinical service that cared for the patient. All of these nuances become integral to optimizing the data collection protocol and are best captured in a coding manual developed specifically for each data collection tool. The coding manual must be pilot tested and will often require revisions. In addition, if multiple sites with unique medical record documentation systems are included in a study, each will likely require a site-specific coding manual. Although it is time-consuming to develop and test, a thoughtful data collection protocol and the use of a well-developed coding manual increase the interrater reliability (IRR) of data abstraction from the medical record (Gearing et al., 2006; Goldman, 1992; VonKoss Krowchuk, Moore, & Richardson, 1995). Clinical staff nurses know where the most reliable data are recorded in the medical record and thus can provide expert consultation to investigators throughout the process of developing, pilot testing, and implementing a coding manual for the purposes of medical record data collection.

3.3. The selection, training, and management of a research team of data abstractors

Optimal data collection from the medical record is contingent on the careful selection and rigorous training of data abstractors (Allison et al., 2000; Gearing et al., 2006; Pan et al., 2005; Wu & Ashton, 1997). Ideally, data abstractors will have the following qualifications: experience with retrospective data collection from the medical record, clinical and research experience relevant to the study being conducted, and advanced educational preparation in a health care profession (i.e., nursing; Gearing et al., 2006). In reality, investigators may not have the option of selecting multiple experienced data abstractors with clinical expertise in the research area of interest and advanced educational preparation. As a result, investigators need to strategize when selecting data abstractors based on who has enough clinical and research experience to understand the work and navigate the medical record. More importantly, investigators will optimize data collection from the medical record by selecting data abstractors who are clinically experienced, interested in research, willing to learn, and able to be trained to follow the data collection protocol. These attributes and abstractor qualifications become significant to investigators aiming to minimize data collection error.

Training and managing data abstractors are time and resource intensive but critical to study success. All data abstractors must complete education and training in the protection of human subjects especially as this training pertains to the protection of health information. In addition, all data abstractors must be added to any study protocol approved by the institutional review board prior to the collection of any data for the purposes of research. Following this foundational research training, data abstractors must be oriented to the study protocol, data collection tool, and the coding manual developed to guide the data collection from the medical record. Data abstractors should remain blinded to the study hypothesis (Gearing et al., 2006; Worster & Haines, 2004), and in case-control study designs, abstractors should remain blinded to which participants are cases and which are healthy controls. The blinding of abstractors enhances the rigor of the study by preventing an individual's bias or interpretation of study data collected from the medical record (Worster & Haines, 2004). Finally, data abstractors must be trained on how to manage missing data and conflicting data found within the medical record. These specific study procedures may be explicit in the coding manual, as well as incorporated into the data management protocol within a study.

Data management throughout any study is important. However, when conducting investigations that involve multiple variables collected from the medical record across

multiple study sites, data often become more complicated to manage. The data management protocol may include the following attributes: scheduled research team meetings that involve all of the data abstractors and the principal investigator (PI) or project director (PD) of the study; regular, ongoing communication with the PI and the PD of the study to resolve questions and data conflicts; and the establishment of an independent committee or panel of experts who serve as consultants to the data abstractors when conflicts arise (Jansen et al., 2005). Of these data management attributes, communication is most important. The use of electronic communication via e-mail, the use of networked shared file areas, and the implementation of Internet-based communication documents (i.e., Google Docs and Google Calendar) have proven effective for prompt communication among team members.

3.4. Ensuring and reporting IRR

Research studies that involve abstraction of patients' medical records rely on the assumption that necessary data will be present in the record, that data will be in a form adequate for abstraction, that data are accurate and consistent throughout the record, and that data will be interpreted in a consistent manner by all the abstractors involved in the study (Eder et al., 2005). The strategies outlined in this article and summarized in Table 1 will help ensure IRR when abstracting data from the medical record. In addition, the following focus points may be followed to enhance IRR:

1. A literature review with a focus on the sensitivity and specificity of data to be collected should be conducted. This will validate that a retrospective medical record review is the most effective method to be used for the study purpose and that the data planned to be collected will adequately answer the research question.
2. If permitted by the study design, the investigator should conduct a pilot study on a smaller sample size, which would be statistically representative of the final sample of participants. Pilot studies maximize IRR in that they are helpful in testing the research instruments and methods and allow the abstractor team to become accustomed to the data collection tools and procedures (Pronovost et al., 2003; Yawn & Wollan, 2005).
3. To increase IRR and to reduce error, the investigator should be prepared to involve a team of two to four abstractors and to conduct periodic audits for accuracy and reliability of data input (Gearing et al., 2006). These auditing exercises to ensure IRR should be planned at multiple time points over the course of the study.
4. The PI should plan to meet routinely with abstractors to discuss the data collection effort and clarify any questions about data being collected. In addition, the investigator should be prepared to revise the data collection tool and refine the coding manual as necessary (New York State Department of Health AIDS Institute, 2008).

Low IRR indicated by discrepancies in data collected by abstractors from patient medical records is a significant limitation associated with retrospective chart review studies and can jeopardize the credibility of the research results. Nevertheless, it was found that most retrospective chart review studies do not discuss IRR limitations or the methods implemented to mitigate these limitations (Allison et al., 2000; Yawn & Wollan, 2005). Thus, when conducting research involving data collection from the medical record, investigators are wise to promote the highest level of IRR through continuous monitoring and periodic reviews, provide a detailed report of limitations encountered during various phases of the study, and discuss how each of those limitations was addressed. By paying special consideration to strategic methods and IRR, the investigator will strengthen the validity of the methods implemented and, in turn, the credibility of study results generated from medical record data (Scanlon et al., 2006; Yawn & Wollan, 2005).

4. Conclusion

The patient medical record has long been recognized as a rich source of information for conducting clinical research. However, investigators must take a strategic approach to data collection efforts and implement a rigorous methodology when conducting prospective and retrospective clinical studies that utilize the data available in the medical record (Faster Cures: The Center for Accelerated Medical Solutions, 2004). A strategic approach to conducting retrospective chart reviews in clinical nursing research will yield high-quality data and accurate results. The strategies suggested in this article include the development and testing of a data collection tool; the use of a coding manual guiding data collection of specific variables from the medical record; and the selection, training, and management of a research team of data abstractors. Careful reporting of the research methodology pertaining to the abstraction and analysis of medical record data, IRR, and study limitations are critical steps in validating outcomes of the study.

References

- Allison JJ, Wall TC, Spettell CM, Calhoun J, Fargason CA Jr, Kobylinski RW, et al. The art and science of chart review. *Joint Commission Journal on Quality Improvement*. 2000; 26(3):115–136. [PubMed: 10709146]
- Bates MC, Rashid M, Campbell JE, Stone PA, Broce M, Lavigne PS. Factors influencing the need for target vessel revascularization after renal artery stenting. *Journal of Endovascular Therapy*. 2006; 13(5):569–577. [PubMed: 17042665]
- Battle CL, Zlotnick C, Miller IW, Pearlstein T, Howard M. Clinical characteristics of perinatal psychiatric patients: A chart review study. *Journal of Nervous and Mental Disease*. 2006; 194(5):369–377. [PubMed: 16699387]
- Bernard, M.; Lida, B.; Riley, S.; Hackler, T.; Janzen, K. A comparison of popular online fonts: Which size and type is best?. *Usability News*. 2002. Retrieved January 18, 2010 from <http://www.surl.org/usabilitynews/41/onlinetext.asp>
- Biederman J, McDonnell MA, Wozniak J, Spencer T, Aleardi M, Falzone R, et al. Aripiprazole in the treatment of pediatric bipolar disorder: A systematic chart review. *CNS Spectrums*. 2005; 10(2):141–148. [PubMed: 15685125]
- Burt, CW.; Hing, E.; Woodwell, D. Electronic medical record use by office-based physicians: United States, 2005. 2006. Retrieved May 6, 2009 from <http://www.cdc.gov/nchs/products/pubs/pubd/hestats/electronic/electronic.htm>
- Cassidy LD, Marsh GM, Holleran MK, Ruhl LS. Methodology to improve data quality from chart review in the managed care setting. *American Journal of Managed Care*. 2002; 8(9):787–793. [PubMed: 12234019]
- Colón-Emeric C, Schenck A, Gorospe J, McArdle J, Dobson L, DePorter C, et al. Translating evidence-based falls prevention into clinical practice in nursing facilities: Results and lessons from a quality improvement collaborative. *Journal of American Geriatric Society*. 2006; 54(9):1414–1418.
- Coyne CA, Xu R, Raish P, Plomer K, Dignan M, Wenzel LB, et al. Randomized, controlled trial of an easy-to-read informed consent statement for clinical trial participation: A study of the Eastern Cooperative Oncology Group. *Journal of Clinical Oncology*. 2003; 21(5):836–842. [PubMed: 12610182]
- DeMille D, Deming P, Lupinacci P, Jacobs LA. The effect of the neutropenic diet in the outpatient setting: A pilot study. *Oncology Nursing Forum*. 2006; 33(2):337–343. [PubMed: 16518449]
- Eder C, Fullerton J, Benroth R, Lindsay SP. Pragmatic strategies that enhance the reliability of data abstracted from medical records. *Applied Nursing Research*. 2005; 18(1):50–54. [PubMed: 15812736]
- Fanaroff JM, Wilson-Costello DE, Newman NS, Montpetite MM, Fanaroff AA. Treated hypotension is associated with neonatal morbidity and hearing loss in extremely low birth weight infants. *Pediatrics*. 2006; 117(4):1131–1135. [PubMed: 16585307]

- Faster Cures: The Center for Accelerated Medical Solutions. Using electronic medical records to bridge patient care and research. 2004. Retrieved May 6, 2009 from http://www.fastercures.org/objects/pdfs/white_papers/emr_whitepaper_summary.pdf
- Gaudreau J, Gagnon P, Roy M, Harel F, Tremblay A. Opioid medications and longitudinal risk of delirium in hospitalized cancer patients. *Cancer*. 2007; 109(11):2365–2373. [PubMed: 17469164]
- Gearing RE, Mian IA, Barber J, Ickowicz A. A methodology for conducting retrospective chart review research in child and adolescent psychiatry. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*. 2006; 15(3):126–134. [PubMed: 18392182]
- Gilbert EH, Lowenstein SR, Kozoi-McLain J, Barta DC, Steinder J. Chart reviews in emergency medicine research: Where are the methods? *Annals of Emergency Medicine*. 1996; 27(3):305–308. [PubMed: 8599488]
- Goldman RL. The reliability of peer assessments of quality of care. *Journal of the American Medical Association*. 1992; 267:958–960. [PubMed: 1734109]
- Gregory KE. Necrotizing enterocolitis: Findings from a retrospective medical record review. *Newborn and Infant Nursing Reviews*. 2007; 7(3):143–150.
- Gunn, PW.; Hansen, ML.; Kaelber, DC. Underdiagnosis of pediatric hypertension—an example of a new era of clinical research enabled by electronic medical records. *Annual Symposium Proceedings/AMIA Symposium*; 2007. p. 966
- Haley RW, Schaberg DR, Mcclish DK, Quade D, Crossley KB, Culver DH, et al. The accuracy of retrospective chart review in measuring nosocomial infection rates: Results of validation studies in pilot hospitals. *American Journal of Epidemiology*. 1980; 111(5):516–533. [PubMed: 7377196]
- Halm EA, Tuhim S, Wang JJ, Rockman C, Riles TS, Chassin M. Risk factors for perioperative death and stroke after carotid endarterectomy. *Stroke*. 2009; 40:221–229. [PubMed: 18948605]
- Jansen A, Aalst-Cohen ES, Hutten BA, Büller HR, Kastelein JJ, Prins MH. Guidelines were developed for data collection from medical records for use in retrospective analyses. *Journal of Clinical Epidemiology*. 2005; 58(3):269–274. [PubMed: 15718116]
- Johnson M, Krosnick A, Carson P, McDade AM, Laraway K. A retrospective chart review of uncontrolled use of metformin as an add-on therapy in type 2 diabetes. *Clinical Therapeutics*. 1998; 20(4):691–698. [PubMed: 9737829]
- Jones K, Fennie K. Factors influencing pressure ulcer healing in adults over 50: An exploratory study. *Journal of the American Medical Directors Association*. 2007; 8(6):378–387. [PubMed: 17619036]
- Murphy EC, Ferris FL, O'Donnell WR. An electronic medical records system for clinical research and the EMR–EDC interface. *Investigative Ophthalmology and Visual Science*. 2008; 48:4383–4389. [PubMed: 17898254]
- Murray MD, Smith FE, Fox J, Teal EY, Kesterson JG, Stiffler T, et al. Structure, functions, and activities of a research support informatics section. *Journal of the American Medical Informatics Association*. 2003; 10(4):389–398. [PubMed: 12668695]
- New York State Department of Health AIDS Institute. NQC quality academy: Collecting performance data. National Quality Center: Improving HIV care. 2008. Retrieved May 9, 2009 from <http://nationalqualitycenter.org/index.cfm/6127/15395>
- Pan L, Fergusson D, Schweitzer I, Hebert PC. Ensuring high accuracy of data abstracted from patient charts: The use of a standardized medical record as a training tool. *Journal of Clinical Epidemiology*. 2005; 58(9):918–923. [PubMed: 16085195]
- Pronovost P, Berenholtz SM, Ngo K, McDowell M, Holzmueller C, Haraden C, et al. Developing and pilot testing quality indicators in the intensive care unit. *Journal of Critical Care*. 2003; 18(3):145–155. [PubMed: 14595567]
- Scanlon MC, Miller M, Harris JM, Schulz K, Sedman A. Targeted chart review of pediatric patient safety events identified by the agency for healthcare research and quality's patient safety indicators methodology. *Journal of Patient Safety*. 2006; 2(4):191–197.
- Sharek PJ, Horbar JD, Mason W, Bisarya H, Thurm CW, Suresh G, et al. Adverse events in the neonatal intensive care unit: Development, testing, and findings of an NICU-focused trigger tool to identify harm in North American NICUs. *Pediatrics*. 2006; 118:1332–1340. [PubMed: 17015521]

- Staller JA. Intramuscular ziprasidone in youth: A retrospective chart review. *Journal of Child and Adolescent Psychopharmacology*. 2004; 14(4):590–592. [PubMed: 15662151]
- Taylor ED, Theim KR, Mirch MC, Ghorbani S, Tanofsky-Kraff M, Adler-Wailes D, et al. Orthopedic complications of overweight in children and adolescents. *Pediatrics*. 2006; 117(6):2167–2174. [PubMed: 16740861]
- VonKoss Krowchuk H, Moore ML, Richardson L. Using health records as sources of data for research. *Journal of Nursing Measurements*. 1995; 3(1):3–12.
- Worster A, Haines T. Advanced statistics: Understanding medical record review (MRR) studies. *Academic Emergency Medicine*. 2004; 11(2):187–192. [PubMed: 14759964]
- Wu L, Ashton CM. Chart review: A need for reappraisal. *Evaluation and Health Professions*. 1997; 20(2):145–163.
- Yawn BP, Wollan P. Interrater reliability: Completing the methods description in medical records review studies. *American Journal of Epidemiology*. 2005; 161(10):974–977. [PubMed: 15870162]

STUDY ID _____
 DATA ABTRACTOR _____
 DATE _____
 DATA ENTERED _____

STUDY TITLE
PI NAME
CLINICAL DATA COLLECTION FORM
FORM VERSION: DATE APPROVED

1. Study ID _____
2. Subject medical record number _____
3. Subject name (last, first) _____
4. Subject date of birth (mm/dd/yyyy) ____/____/____

RELEVANT DEMOGRAPHIC DATA

[Demographic variables and data points will vary based on individual attributes of the study]

5. Gender _____
 Male (0)
 Female (1)
6. Race/Ethnicity _____
 White, non-Hispanic (1)
 White, Hispanic (2)
 Black, non-Hispanic (3)
 Black, Hispanic (4)
 Asian or Pac. Islander (5)
 Native American (6)
 Other (7)
7. Years of education _____
 0-8 years (1)
 9-12 years: high school or some high school (2)
 12-16 years: college or some college (3)
 >16 years: post college education (4)

CLINICAL DATA POINTS

[Study variables and data points determined by investigator, corresponding to study framework]

8. Date of diagnosis of disease (mm/dd/yyyy) ____/____/____
9. Weight at diagnosis (Kg) _____
10. BMI at diagnosis _____
11. Systolic blood pressure at diagnosis _____
12. Fasting blood sugar at diagnosis _____

STUDY ID _____
 DATA ABTRACTOR _____
 DATE _____

LABORATORY VALUES- OBTAINED AT TIME OF DIAGNOSIS

[Study variables, data points, and time at which collected determined by investigator]

13. WBC with Differential (cells/cubic millimeter (cmm)) _____
14. RBC (cells/cmm) _____
15. Hemoglobin (HgB) (gms/deciliter) _____
16. Hematocrit (Hct) (%) _____
17. Platelets (Plt) _____
18. Sodium (Na) _____
19. Potassium (K) _____
20. Chloride (Cl) _____
21. Bicarbonate (or total CO2) _____

Fig. 1.

Sample data collection form.

**CODE BOOK
STUDY TITLE (PI NAME)
CODE BOOK VERSION: DATE APPROVED**

1. Study ID
Definition: ID assigned solely for purposes of study by PI at time of enrollment
Best source: Study enrollment log; found in locked file-cabinet of study office
2. Subject medical record number
Definition: Hospital medical record number
Best source: Stamped faceplate on medical record pages
3. Subject name (last, first)
Definition: Last and first name under which subject admitted to hospital
Best source: Stamped faceplate on medical record pages
4. Subject date of birth (mm/dd/yyyy)
Definition: Self explanatory
Best source: Stamped faceplate on medical record pages

RELEVANT DEMOGRAPHIC DATA

5. Gender
Definition: Self explanatory
Best source: Attending physician admission note OR medical record demographic data sheet
6. Race/Ethnicity
Definition: Self-identified race and ethnicity reported by subject
Best source: Attending physician admission note OR medical record demographic data sheet
7. Years of education
Definition: Years of education as reported by subject
Best source: Attending physician admission note OR medical record demographic data sheet

CLINICAL DATA POINTS

8. Date of diagnosis of disease (mm/dd/yyyy)
Definition: The date on which the subject was first diagnosed with the disease of study
Best source: History and physical documentation OR attending physician admission note
9. Weight at diagnosis (Kg)
Definition: Subject body weight, in kilograms, at the time of diagnosis of disease of study
Best source: History and physical documentation OR attending physician admission note
10. BMI at diagnosis
Definition: Body Mass Index at disease diagnosis
Best source: History and physical documentation OR attending physician admission note. If NOT documented, calculate with following formula: $BMI = \frac{(\text{weight in pounds}) \cdot 703}{(\text{height in inches})^2}$
11. Systolic blood pressure at diagnosis
Definition: The highest blood pressure measured in the arteries during systole
Best source: History and physical documentation OR attending physician admission note
12. Fasting blood sugar at diagnosis
Definition: Level of glucose present in the blood drawn at a minimum of eight hours after fasting
Best source: Attending physician admission note, history and physical documentation, or laboratory report from morning following admission [use hospital lab on-line reporting system]

LABORATORY VALUES- OBTAINED AT TIME OF DIAGNOSIS

[NOTE: If labs not documented in history and physical, see laboratory report in hospital lab on-line reporting system, documenting labs drawn on morning following admission]

13. WBC with Differential (cells/cubic millimeter (cmm))

Definition: Total WBC count including number of circulating neutrophils, lymphocytes, monocytes, eosinophils and basophils. Part of the complete blood count (CBC)

Best source: Lab report obtained on morning following admission

14. RBC (cells/cmm)

Definition: Number of circulating red blood cells

Best source: Lab report obtained on morning following admission

15. Hemoglobin (Hgb) (gms/deciliter)

Definition: Iron-containing oxygen-transport protein in the circulating red blood cells

Best source: Lab report obtained on morning following admission

16. Hematocrit (Hct) (%)

Definition: The percentage of red blood cells in a volume of whole blood

Best source: Lab report obtained on morning following admission

17. Platelets (Plt)

Definition: Type of blood cell that helps blood to clot, measured in number of thousands

Best source: Lab report obtained on morning following admission

18. Sodium (Na)

Definition: Electrolyte measured in the blood

Best source: Lab report obtained on morning following admission

19. Potassium (K)

Definition: Electrolyte measured in the blood

Best source: Lab report obtained on morning following admission

20. Chloride (Cl)

Definition: Electrolyte measured in the blood

Best source: Lab report obtained on morning following admission

21. Bicarbonate (or total CO₂)

Definition: An acid carbonate; one of the buffers necessary to maintain normal acid-base balance

Best source: Lab report obtained on morning following admission

Fig. 2.

Sample code book for data collection.

Table 1

Strategies for conducting medical record review in clinical nursing research

Strategy	Specific research actions
Development and testing of a data collection tool	<ul style="list-style-type: none"> • Define research aims and identify study variables. • Collaborate with clinical staff to assess whether necessary data are available in the medical record, and if so, where the best source of the data is located within the record. • Design the data collection tool such that the variables to be recorded on the tool correspond to the data as it appears in the medical record. • Design the data collection tool in an easy-to-use fashion, with the use of charts, tables, and easy-to-read fonts and formatting. • Include the study ID number on each page of the data collection tool. • Pilot test the data collection tool with a subset of medical records.
Use of a coding manual	<ul style="list-style-type: none"> • List each variable with an explanation as to how the variable should be recorded on the data collection tool. • Provide the best source of data within the medical record for each study variable. • Pilot test and revise the coding manual as appropriate. • Consult with clinical nurses throughout the process of developing, pilot testing, and implementing the coding manual.
The selection, training, and management of data abstractors	<ul style="list-style-type: none"> • Select data abstractors who are clinically experienced, trained in research, and able to explicitly follow the study protocol. • Ensure that all data abstractors complete human subjects protections training and understand the importance of keeping protected patient information confidential and secure. • Implement regular and ongoing communication with all data abstractors via electronic means and in-person meetings. • Keep data abstractors blind to case and control or treatment and intervention subject assignment whenever possible. • Establish an independent panel of experts who will serve as consultants to the data abstractors when conflicts arise.