A practical guide to conducting and writing medical record review studies

Asad E. Patanwala, Pharm.D., M.P.H., BCPS, FCCP, FASHP, Department of Pharmacy Practice and Science, College of Pharmacy, University of Arizona, Tucson, AZ.

Purpose. A practical guide to conducting and reporting medical record review studies, including instruction and insights on topics ranging from idea conception to manuscript submission, is provided.

Summary. Research based on the retrospective collection and analysis of data from medical records is often the first type of research conducted by new investigators; such research can be particularly useful in high-acuity populations and is often more feasible and cost-effective than prospective studies. The research process begins with formulating a valid hypothesis-driven or exploratory research question and refining it into one that is testable; this can be facilitated by incorporating information from various sources (e.g., relevant literature, consultation with experts, observations of practice). Accurate sample size calculations are important in ensuring study feasibility (i.e., the estimated sample sizes required can be obtained). Appropriate data collection methods are particularly crucial in medical record review studies; key considerations include pertinent variables, systematic subject identification, standardized data collection instruments, training of data collectors to minimize interrater variability, processes to ensure the accuracy and reliability of data, and blinded collection of the outcome variable. Checklists are provided to guide investigators in retrospective data collection, manuscript preparation, and decisions regarding authorship eligibility and sequencing.

Conclusion. The quality of medical record review studies and the likelihood of their acceptance for publication can be improved through adherence to recommended standards of research design, data collection and analysis, ethical authorship, and manuscript preparation.

Keywords: manuscripts, medical manuscripts, publishing, retrospective studies, writing

Am J Health-Syst Pharm. 2017; 74:1853-64

Medical record review studies are a subset of observational investigations that are very common in clinical research.[1] These studies may have cohort, case–control, or cross-sectional designs.[2] Conducting these types of projects is popular, because they are relatively feasible and can be completed with fewer resources than are required for other types of research. For these reasons, new practitioners usually embark on their first research experience by using data from medical records.[3] Such studies can be particularly useful in some fields of research involving high-acuity patient populations, where substantial barriers to conducting prospective studies exist.[4] Clinical trials may require considerable resources or may not be feasible because of the nature of the disease, consent process, environment, or timing of therapies. Given these obstacles, investigators are able to rely on the natural variability in practice (e.g., provider prescribing preferences) or deliberate practice changes (e.g., formulary changes) to make retrospective comparisons.

This article is part of a special series on research and publishing.

Address correspondence to Dr. Patanwala (patanwala@pharmacy.arizona.edu).

DOI 10.2146/ajhp170183

The Oxford Center for Evidence-Based Medicine considers observational research as lower-level evidence.[5] Nonetheless, high-quality observational studies have an important role in the literature because they are hypothesis generating and can serve as preliminary evidence for future trials. Furthermore, well-designed observational research can provide treatment effect estimates that are consistent with those from randomized controlled trials[6,7] and are a critical component of comparative effectiveness research.[8-10] The increasing availability and quality of data in electronic medical records in health systems have also improved our ability to extract and analyze information to answer important clinical questions more efficiently. Thus, there are great opportunities to make valuable contributions to the literature via observational research that includes medical record review. *AJHP* has published a series of articles that serve as a toolkit for practice-based research.[11-23] The articles in this series provide detailed information that includes the entire scope of conducting research. Also, investigators interested in database analyses are referred to guidance provided by the International Society for Pharmacoeconomics and Outcomes Research.[24] Information regarding the use of automated data sets is beyond the scope of this article. However, there remains a need for a concise document that can be used as a reference to guide investigators conducting medical record review studies from idea conception to manuscript preparation, which will help optimize project success and final publication.

The purpose of this article is to provide investigators with a practical guide to conducting and writing medical record review studies involving manual data collection for both clinical and nonclinical research questions. It is important to emphasize that the information provided in this article is no substitute for consultation with experienced researchers early in the process of project planning. The text of this guide has been intentionally minimized with a preference for tables, figures, and checklists that can be used as a quick reference. Examples have been included throughout to facilitate application of the content. Finally, guidance for writing manuscripts related to studies involving medical record review is provided.

KEY POINTS

- Retrospective studies should begin with a well-built research question, feasibility assessment, and sample size calculation.
- Vigorous data collection methods should be used to minimize bias and maximize the reliability of the data.
- The use of checklists to facilitate appropriate data collection, evaluation of authorship eligibility, and manuscript preparation can improve study quality and reporting of research results.

Refining the study question

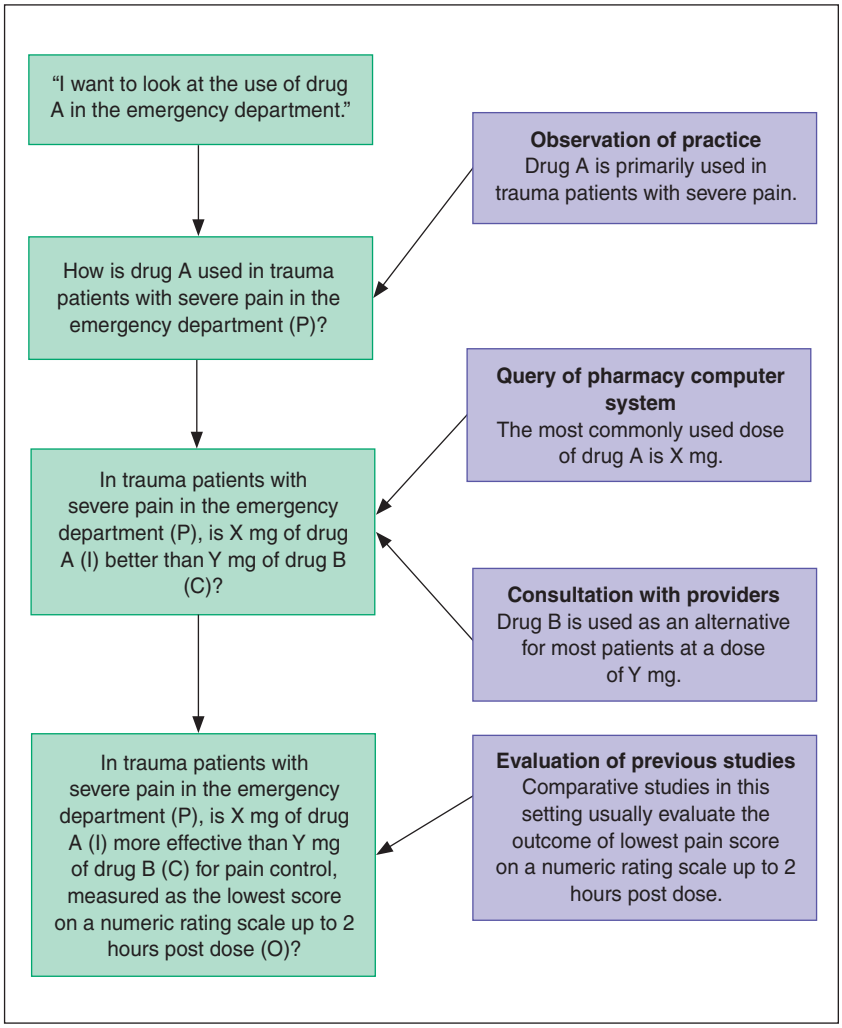
Identifying and developing a good research question are prerequisites to any successful project.[15] The FINER criteria (feasible, interesting, novel, ethical, relevant) are a good starting point to determine if a question is worth pursuing.[25] However, initial research questions tend to be broad and may not go beyond identifying an area of practice that is of interest to the investigator. Ideally, research questions need to be well built and have great precision. These goals can be accomplished by structuring the question in the PICO (patient or population, intervention, comparator, outcome) format.[26] Figure 1 depicts a theoretical example of a study idea being refined into a research question. The initial idea seems to be rather vague. In fact, the initial question is not really a question but rather a domain of practice (i.e., use of drug A in the emergency department). The question is improved and has a greater degree of precision after evaluation of the existing literature, consultation with experts, and observations of practice. In theory it could be possible to refine the question further by narrowing the patient population (e.g., to patients with a specific type of injury), but as the population becomes more targeted and homogeneous, the feasibility of the project with regard to sample acquisition is decreased. Thus, although formulating a well-built PICO question is important, it may not necessarily translate into the ability to pursue a given project. Investigators must understand this balance and go through an iterative process that cycles between the research question and project feasibility. Also, not all research questions for medical record review studies fit well into the PICO format. Some investigations may just be descriptive evaluations without an intervention or comparison (e.g., a study of the rate of adverse events in a certain setting). Some investigations may be more exploratory than hypothesis driven. However, the process of thinking through PICO can help define and refine the components applicable to a given study.

Feasibility assessment and sample size

Obtaining patient lists. One of the most common threats to project success with medical record reviews is the inability to acquire the necessary sample. After construction of the PICO question, it is imperative for the investigator to have access to appropriate data query tools to obtain the target sample. Most pharmacy systems can identify potential subjects on the basis of medications dispensed or administered. However, this list of subjects is usually not restricted to the patient population of interest, and it requires a substantial manual effort to drill down to the subjects who may be eligible for inclusion. As the numbers of inclusion and exclusion criteria in-

crease, the investigator will have to sift through a greater number of records to obtain the necessary sample, because the number of patients eligible for study entry will decrease as criteria become more stringent; unfortunately, this seems to be the extent to which some pharmacy departments are able to query the medical record electronically. Given the resources needed to manually identify study subjects within a patient list, projects are often initiated with this limited information. Pursuing a project without a high level of certainty that a sufficient sample can ultimately be obtained is a risky endeavor and increases the chances of project failure. Access to patient lists based on *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*) codes can be useful in identifying patients with certain disease states and procedures, but this process can also be limited by the accuracy of *ICD-9-CM* coding. Ideally, systems would have the ability to query electronic medical records using a combination of *ICD-9-CM* codes, demographics, medication-use data, and other clinical information; such capability is increasingly available within health systems. If possible, investigators should merge lists based on *ICD-9-CM* codes with those based on medication lists as a starting point to determine if the study is feasible. If the sample is inadequate, then the PICO question should be revised and the process repeated, as described in the previous section.

Figure 1. Process of constructing a research question in the PICO (patient, intervention, comparison, outcome) format. As more information is collected, increasingly refined questions addressing PICO elements are formulated until a final research question addressing all 4 elements is determined.



Sample size determination. A common misconception is that sample size calculations are not necessary when conducting medical record review studies. Indeed, one of the major criticisms encountered during the publication process is that a sample size determination was not completed and, thus, the study is susceptible to Type II error (i.e., the study did not find a significant difference when one truly existed). Due to the effort needed for manual data collection, many medical record review studies tend to have relatively small samples. Thus, it is good practice to conduct an a priori sample size determination. It is highly recommended that investigators consult with a biostatistician early in the research process.[21] However, it may be difficult to have such statistical expertise readily available at many institutions, especially when study feasibility is being determined. Tables 1 and 2 provide sample size estimates for comparisons based on common types of outcome variables (continuous and categorical) for 2 independent groups. In the absence of a biostatistician, these tables may be used and referenced for estimates of sample size in order to determine project feasibility during formulation of the research question. If the desired sample size cannot be obtained, then investigators should reevaluate the feasibility of conducting the project or make necessary adjustments to their study design. The hypothetical examples presented below provide an interpretation of the concepts outlined in Tables 1 and 2 to assist investigators in applying those concepts in the study design process. A power of 80%, a 2-sided alpha of 0.05, and

an allocation ratio of 1:1 have been used in all examples. All calculations in Tables 1 and 2 were conducted in Stata, version 13 (StataCorp LLC, College Station, TX). There are many caveats pertaining to various aspects of sample size calculations, such as the allocation ratio between groups, the number of groups, and comparisons of repeated measures. A change in any of these parameters would change the sample size. These variations are not described in this paper. Instead the reader is provided with 2 examples that follow the PICO format, which together provide an overview of the process of sample size estimation.

Table 1. Sample Size Calculation for Comparison of 2 Independent Means[a]

Mean 1 – Mean 2 S.D.	*n*[b]
0.10	1,571
0.15	699
0.20	394
0.25	253
0.30	176
0.35	130
0.40	100
0.45	79
0.50	64
0.55	53
0.60	45
0.65	39
0.70	34

[a]Calculations assume a power of 80%, 2-sided alpha of 0.05, and same S.D. value in both groups.
[b]*n* denotes necessary sample size in each group.

Example 1: Comparing means in 2 independent samples. Consider the following PICO-formatted research question: In adult trauma patients requiring mechanical ventilation (P), is drug A (I) more effective than drug B (C) with regard to ventilator-free hours measured at 7 days (O)? Based on published data from a placebo-controlled study in which the investigators tested the hypothesis that use of dexmedetomidine in addition to standard care would result in a 20-hour between-group difference in mean ventilator-free hours over a 7-day period,[27] we decide to test for a 20-hour difference as well. In our example, let us assume that the mean estimated values for ventilator-free hours are 140 hours with the use of drug A (mean 1) and 120 hours with the use of drug B (mean 2); also, for simplicity we assume a common S.D. value of 40 hours in both groups. Thus, according to the formula shown in Table 1 (|mean 1 – mean 2 |/S.D.), we calculate as follows: (|140 – 120 |/40 = 0.5. Based on these assumptions and calculations, we estimate that a sample of 64 patients is required in each group (a total sample of 128).

Example 2: Comparing proportions in 2 independent samples. Consider the following PICO question: In adult trauma patients with brain injury (P), is drug A (I) more effective than drug B (C) in terms of in-hospital mortality (O)? Determining what is a clinically meaningful difference in an outcome such as mortality is often based on the clinical context, previous studies, and the judgment of the investigators. In this scenario, we will test for a 15% absolute difference in in-hospital mortality. Let us assume mortality rates of 30% with drug A and 15% with drug B. The effect size can be calculated, according to the formula shown in Table 2, as follows: |proportion 1 – proportion 2 |, with those 2 values denoting the propor-

Table 2. Sample Size Calculation for Comparison of 2 Independent Proportions[a,b]

	Effect Size (P1 – P2)											
	5%		10%		15%		20%		25%		30%	
P1 (%)	P2 (%)	*n*[c]	P2 (%)	*n*[c]	P2 (%)	*n*[c]	P2 (%)	*n*[c]	P2 (%)	*n*[c]	P2 (%)	*n*[c]
90	85	686	80	199	75	100	70	62	65	43	60	32
80	75	1,094	70	294	65	138	60	82	55	54	50	39
70	65	1,377	60	356	55	163	50	93	45	61	40	42
60	55	1,534	50	388	45	173	40	97	35	62	30	42
50	45	1,565	40	388	35	170	30	93	25	58	20	39
40	35	1,471	30	356	25	152	20	82	15	49	10	32
30	25	1,251	20	294	15	121	10	62	5	36	. . .[d]	. . .
20	15	906	10	199	5	76	. . .	. . .	. . .	. . .	. . .	. . .
10	5	435	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .

[a]P1 = proportion 1, P2 = proportion 2.
[b]Assuming a power of 80% and a 2-sided alpha of 0.05.
[c]*n* denotes necessary sample size in each group.
[d]Not applicable.

tions of patients who died with the use of drugs A and B, respectively (i.e., 30% – 15% = 15%). Based on these assumptions, we can see in Table 2 that 121 patients are required in each group (a total sample of 242).

These examples highlight some of the information investigators will need (and a biostatistician would request) for a formal sample size calculation. A search should be conducted for prior studies with similar designs from which requisite items of information (e.g., proportions, means, S.D. values) can be obtained for the outcome of interest. If no similar studies exist, assumptions can be made on the basis of data from studies of similar populations in which these variables have been measured. The research team should also come to consensus regarding the effect size (i.e., the difference between groups) that is considered to be clinically meaningful; ideally, this information can be obtained from reports on previous studies, as in example 1, but if none is available, decisions can be based on clinical expertise. If a larger effect size is chosen (as in example 2), this should be considered when interpreting the results of the study. For instance, let us say that in example 2 above, the in-hospital mortality rates ended up being 20% with drug A and 30% with drug B ($p = 0.123$). In this scenario, although the difference between groups was not statistically significant, the lack of significance may have been a function of what the investigators considered an important difference (in example 2 above, that was an absolute difference of 15%); in other words, the study was simply not powered to test for a 10% absolute difference in mortality—a difference that may certainly be considered clinically meaningful. That type of information should be acknowledged in the limitations section of the research manuscript.

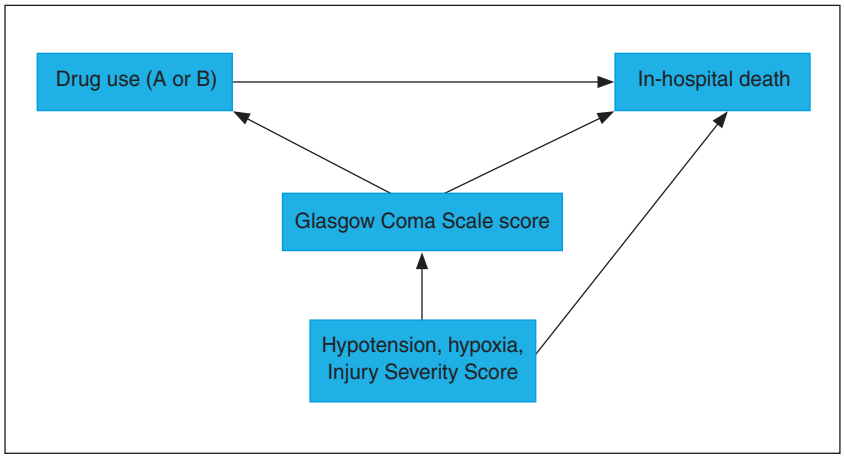
The tables provided in this article should not be construed to obviate consultation with a biostatistician. However, they provide reasonable estimates that can be used while considering certain research questions before embarking on a project. Investigators may also consider using free tools such as G*Power (Heinrich Heine University Düsseldorf, Düsseldorf, Germany), depending on their comfort level with conducting sample size calculations.[28]

Study variables and data collection

There are some issues related to data collection that are unique to medical record review studies. The process by which the data are obtained should be optimal, and integrity in data acquisition ensures that study results are reliable. Poor planning can have great implications with regard to the results. As an example, let us consider 2 data collectors obtaining information from the medical record regarding the rate of occurrence of adverse events. This variable is subject to interpretation, and without clear guidance there could be great variability between data collectors in terms of the data collection process. Thus, processes need to be established to test for and mitigate these problems. Below are some important components of the data collection process that should be incorporated into study design. Appendix A is a checklist related to data collection that is provided for the convenience of investigators.

Pertinent study variables. Determinations regarding which study variables to collect should be given considerable thought. Previous studies that investigated outcomes similar or related to the PICO question of interest are an appropriate starting point. It is good practice for the investigative team to develop a list of variables and theoretical causal pathways via brainstorming. The use of causal diagrams in epidemiologic research has been suggested.[29-31] Causal diagrams connect potential causes and effects via arrows. Although a detailed explanation of the theory related to causal diagrams is beyond the scope of this article, the diagrams are intuitive and can help researchers identify important variables that should be collected. Figure 2 is a hypothetical causal diagram related to example 2 above, which could be useful in deter-

Figure 2. Theoretical causal pathway depicting potential variables of interest for a study of the effect of using drug A versus drug B (the cause) on in-hospital mortality (the outcome) in trauma patients with brain injury. As depicted with arrows, scores on the Glasgow Coma Scale, a widely used measure of brain injury severity, can affect both drug selection and mortality. Hypotension, hypoxia, and Injury Severity Score can affect Glasgow Coma Scale scores and are also independently associated with mortality.



mining variables to be collected in a study of the effects of treatment with drug A versus drug B in patients with brain injury. This diagram could probably be developed further to elucidate various relationships pertaining to the research question.

One critical flaw that may be encountered during the peer-review process is related to important study variables that were never collected. Due to this pitfall, some investigators err on the side of collecting excessive information, which makes the data collection process overly burdensome. The right balance needs to be achieved to determine which variables are integral to the study versus those that are just "nice to obtain." In fact, the ethical approval of studies requires that investigators have a reason for every piece of information collected; this minimizes risks related to privacy and breach of confidentiality. In example 2, collecting the Glasgow Coma Scale (GCS) score (a measure of mental status widely used in trauma and intensive care populations) would be critical, as indicated by our causal diagram; if this variable is not available in the medical record, the study should not be pursued. Alternatively, if the study is to focus on patients with long-bone injuries rather than head injuries, the lack of availability of GCS scores may not be a crucial factor in conducting the study.

If continuous or ordinal variables are to be categorized for analysis, then this should ideally be determined a priori. As an example, patients may be categorized by the variable of age into 2 groups (<65 years of age and ≥65 years of age). In some cases, it may not be possible to know the distribution of variables ahead of time, but categorization may still be required because the data are very skewed. Investigators need to be transparent in defining and reporting these variables. Ideally, there should be precedence for categorizations rather than creation of arbitrary groups. For instance, in example 2 above, GCS scores could be categorized into 3 groups according to official definitions: mild (GCS score of 13–15), moderate (GCS score of 9–12), and severe (GCS score of 3–8) brain injury.[32]

Systematic subject identification. Ideally, the final study sample should be representative of the population of interest. The process of identifying patients to be included in the sample can be fraught with bias. In other words, investigators may subconsciously choose to include patient data that would support their hypothesis. Techniques to minimize this bias depend on the study design. As an example, in a cohort study design, patients may be chosen chronologically within a certain range of admission dates, and the reasons for exclusion of patients admitted during that period should be documented. The inclusion and exclusion criteria should be clearly defined and not subject to interpretation. Some journals require investigators to submit a flow diagram depicting this process. Case–control studies are particularly prone to bias during selection of controls. Statistical approaches to matching, as with propensity scores or nearest-neighbor algorithms implemented by computer programs, may help minimize bias from manual selection; unfortunately, the data needed for such matching are needed a priori. Whatever process is used must be clearly described, and a high degree of transparency is vital.

Standardized data collection instrument. Variables should be collected in a manner that makes them conducive to data analysis. Each variable should be categorized or be in numeric form. One rule of thumb is that the data collection form should not contain any free text. For example, the variable "comorbidities" might be listed on the data collection sheet in either of the following formats:

(1) Comorbidities: ☐ diabetes ☐ hypertension ☐ other
(2) Comorbidities: _____

The latter format solicits a free-text response that may require a lot of effort to categorize and analyze later; the former is preferred, but its use requires more planning to ensure that categories are assigned for all comorbidities of interest. Web-based data collection applications such as REDCap[33] have greatly improved this process and have built-in functions to help improve the accuracy of data collection (e.g., the data collector is alerted when the number for a field exceeds a specified threshold value). The data can also be directly exported to statistical software, and a codebook automatically created. Before data collection begins, the form should be tested and modified as necessary.

Training of data collectors. After the final data collection form has been developed, all data collectors should be provided with appropriate training regarding how data should be collected and where the desired variables are located in the medical record. In some electronic health records, the same variable may be reported in more than one location (e.g., in physician notes and also in nursing staff documentation, which may not necessarily contain the same information). All data collectors should follow a consistent process, such as obtaining data only from a specific field that was deemed to contain the most accurate data. The investigators should vet each variable field and its location in the medical record. When multiple data collectors are involved, one strategy to help ensure data integrity is for them to collect data together for the first few records and discuss potential problems before performing data collection individually.

Ensuring data accuracy and reliability

Errors during manual data collection are very common. Ideally, a second investigator should independently check the data collected for accuracy. One quality-control strategy that can be used when there are 2 data collectors is for each of them to collect data on 50% of the sample and then check each other's work. Any

discrepancies should be resolved via consensus. Another technique that has been recommended is for a subset (e.g., 10%) of the sample to be checked by a second investigator.[34] Interrater reliability testing should then be performed and reported in the manuscript for pertinent variables, as identified by the investigators.[4,34] However, there remain uncertainties regarding which variables need this testing, and the process can be unwieldy. Interrater reliability testing is certainly necessary for variables that require judgment, such as the presence or absence of an adverse reaction. The final status of the variable for each subject in the entire cohort (e.g., an adverse reaction occurred or did not occur) may be determined via consensus. In the current era of electronic medical records, some variables are clearly listed and very well defined (e.g., in-hospital mortality, length of stay). In these circumstances, interrater reliability testing is less necessary.

Blinded collection of outcome variables. Ideally, the data collectors should be blinded to the study outcome to minimize bias.[4,34] However, this is difficult to accomplish, because most data collectors for projects (e.g., students, residents) are involved in the study design. Data collection alone does not meet criteria for authorship; hence, data collectors are most likely familiar with the study objectives and hypotheses, which makes blinding a considerable challenge. However, it is possible to minimize this potential bias. One strategy is to segregate collection of the outcome variable from the rest of the data (i.e., use a separate form linked only by a study identifier). A different individual can then collect the outcome variable. This process can help minimize bias if the individual collecting the outcome variable cannot readily determine from the medical record which group the patient should be assigned to (e.g., drug A versus drug B). In some situations the main outcome variable can be obtained via automated electronic query methods (e.g., if hospital length of stay is the variable of interest), which can also help minimize the potential for bias.

Data analysis

For the purposes of this discussion, let us assume that a biostatistician will conduct data analyses. However, investigators can take some steps themselves by performing data checks for each of the collected variables. Such checks will make data analysis much easier and faster for the biostatistician. Training in biostatistics is not required to perform data checking, and programs such as Microsoft Excel (Microsoft Corporation, Redmond, WA) may be used. Each variable can be assessed for errors by identifying outliers. For example, documentation indicating a patient weight of 200 kg may be erroneous and may possibly reflect incorrect entry of the patient's weight in kilograms rather than pounds. Similarly, 2 or more variables can be used to identify potentially erroneous data. For example, if a patient died on hospital day 3, then a variable indicating blood transfusion on day 5 is either incorrect or indicative of patient survival past day 3. In observational studies, a multivariable analysis is often needed to adjust for pertinent confounders. As described in the previous section, the investigators should indicate to the biostatistician what confounders should be considered in any model based on causal theory and clinical expertise. While an overview of biostatistics is beyond the scope of this article, investigators are encouraged to review a previously published article pertaining to common statistical tests used for clinical research data.[20]

Authorship

One of the most sensitive issues for the investigative team pertains to authorship eligibility and sequence. In medical record review studies, the task of manual data collection is commonly assigned to trainees or the junior investigators on the team. Trainees may spend numerous hours manually collecting data. However, according to International Committee of Medical Journal Editors (ICMJE) criteria, data collection alone does not meet the requirements for authorship.[35] The ICMJE criteria are listed in Appendix B. These criteria should be discussed early with the team, especially when working with trainees who may be unfamiliar with the authorship process. Another potentially contentious issue is related to authorship sequence. The authorship rank is based on the amount of work done on the project (i.e., the first author listed did the most work). This issue should be addressed upfront, during project planning, and tasks should be assigned accordingly. Appendix B includes a scoring system using components of the ICMJE criteria. Research teams can consider using this rubric as an objective method of determining the sequence of authors. It is possible that the authorship order may be subject to change based on effort during the course of the study. All authors (including the senior author) must be actively engaged. If ICMJE criteria are not met for any author, then he or she may be removed from the author list; this applies across the spectrum, from junior to senior research team members. As an example, a project mentor who only provides cursory input would not meet the ICMJE authorship criteria. Similarly, authorship should not be granted simply because individuals grant investigator access to their data. Although the authorship order is based on the amount of work done, there is an exception: It is common convention for the senior author who oversees all aspects of the research (i.e., the project mentor) to be listed as the last author if he or she desires. At a minimum, it is good practice to determine the first, second, and senior authors at project initiation.[12] Finally, the team should identify a corresponding author who will communicate with journal editors during the publication process and serve as a point person for the article after publication. The corresponding author is usually the first or senior author on the team.

Institutional review board

Misconceptions regarding the role of institutional review board (IRB) oversight of medical record review studies persist. Investigators must obtain IRB approval prior to conducting any data collection. In some situations, studies using existing data may meet 1 or more criteria for exempt status. It is important to remember that exempt status is determined by the IRB rather than the investigator. One factor to consider with regard to qualifying for an exemption is the extent to which research subjects may be individually identified. The collection of certain data elements that are deemed individually identifiable (e.g., protected health information under Health Information Portability and Accountability Act regulations) may prevent an exempt status determination; however, the study may still qualify for expedited IRB review. Consult the IRB for information on the review and approval process. In some circumstances, investigators may consider a project to constitute quality improvement rather than research. However, a designated committee at the institution must make this determination.[23] Investigators should seek information at an institutional level regarding the approval of such projects. In other words, investigators must be prepared to provide a letter of approval or exemption written by the IRB or quality-improvement committee (if applicable) for all projects, if requested by the journal. IRB approval is reported in the methods section of the manuscript.

Manuscript preparation

Components and structure. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations[1] and the "REporting of Studies Conducted using Observational Routinely-collected health Data" (RECORD) statement,[36] which is an extension of the STROBE document, serve as starting points for determining the reporting requirements for manuscripts reporting on medical record review studies. In fact, the recommendations can serve as a good template early in the research planning process; that way, items that need to be reported can be incorporated into the design and logistics. It is important to note that great manuscript writing cannot overcome flaws in the research. As an example, if the investigators did not perform an a priori sample size calculation, then it will be difficult to explain how the study size was arrived at and how the possibility of Type II error was minimized. Similarly, a well-drafted manuscript that does not contain the key variables needed to answer the research question will very likely be rejected during the peer-review process. Thus, the processes of conducting research and preparing the manuscript should not be viewed in isolation.

Previously published articles provide guidance to authors regarding the manuscript preparation process.[37-42] In addition, ICMJE has endorsed the Introduction, Methods, Results, and Discussion (IMRAD) structure for original research studies and has provided recommendations for each of the IMRAD components.[35] These concepts apply to medical record review studies, with additional emphasis placed on certain components such as data collection. Appendix C summarizes some of the key items specific to such studies that authors should consider; it can serve as a checklist when drafting the manuscript. This appendix should be used in conjunction with the RECORD and STROBE statements and ICMJE recommendations.

Writing process. Although a detailed algorithm for writing research manuscripts has been proposed,[39] and articles have been written on this topic,[37-42] there is no universally accepted process that is suitable for all investigators. A few principles that can make the manuscript writing process more efficient are described below. Although this article is focused on medical record review studies, many of these concepts apply to all studies regardless of design.

1. Start by creating the tables and figures. These article elements serve as a useful reference when writing the main text. During this process, decisions can also be made regarding what information is best suited for presentation in the main text versus tables. The tables can be modified as the text is written to minimize overlap in data presentation.
2. It is common for investigators to start writing the main text sequentially from beginning to end—from introduction to methods to results to discussion (i.e., in the IMRAD format). However, it may be more efficient to write the sections in a different sequence methods, results, discussion, introduction (MRDI). The reason for this approach is that what was done (M) and what was found (R) help guide what is written in the other sections (D and I). One common mistake is providing too much detail regarding previous investigations in the introduction. By writing the introduction last, the authors can utilize some of the more detailed information from the discussion to draft the introduction. The introduction should be relatively brief. As a general guide, in most of the major medical journals, the introductions typically range from 300 to 400 words.
3. The discussion is usually the most challenging section to write. The first paragraph should summarize the key findings and implications of the research. It is important to avoid restating the results in detail. Instead, only a broad summary should be provided. The rest of the discussion is not meant to be a thorough review of the literature (i.e., do not write it like a literature review). Instead, elaborate on how the study adds to the literature by highlighting similarities and differences with the available body of evidence that is most pertinent to the current investigation. Provide insight, interpretations, and applications to practice. The last 2 paragraphs usually discuss limitations and conclusions; how-

ever, this may vary based on journal requirements.

4. Write the abstract after the main text is complete. The abstract serves as a summary of the article. Once the main text is written, most of the abstract content can by be taken verbatim from the main text. This process also ensures consistency between the abstract and main text. If the abstract has already been drafted (e.g., a preliminary abstract presented at a meeting), it can be modified based on the main text to incorporate any changes that may have occurred subsequent to preliminary data analyses.
5. Use subheadings in the methods section. Some subheadings have been suggested in Appendix C, but these can be modified based on the study design and the journal's requirements. The use of subheadings in the first draft of the manuscript ensures that the required items are reported. It is also easier for readers to locate the various components of the methods.
6. Ensure that the manuscript adheres to formatting requirements for the journal selected as the writing process begins. This is to avoid late changes due to journal requirements (e.g., abstract word limits), which would require additional reviews by all authors. If the investigators decide to submit the manuscript to a journal other than the initially targeted journal, or if the manuscript was rejected and a new submission to a different journal is being planned, the authors must verify that the newly targeted journal's formatting guidelines are followed.
7. After the manuscript is prepared, obtain feedback on the draft from some experienced colleagues who are not involved in the research. This will help identify potential problems and enable revisions prior to final submission.

Conclusion

The quality of medical record review studies and the likelihood of their acceptance for publication can be improved through adherence to recommended standards of research design, data collection and analysis, ethical authorship, and manuscript preparation.

Disclosures

The author has declared no potential conflicts of interest.

References

1. Von Elm E, Altman DG, Egger M et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol.* 2008; 61:344-9.
2. Mann CJ. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emerg Med J.* 2003; 20:54-60.
3. McKelvey RP, Hatton RC, Kimberlin CA. Pharmacy resident project publication rates and study designs from 1981, 1991, and 2001. *Am J Health-Syst Pharm.* 2010; 67:830-6.
4. Gilbert EH, Lowenstein SR, Koziol-McLain J et al. Chart reviews in emergency medicine research: where are the methods? *Ann Emerg Med.* 1996; 27:305-8.
5. OCEBM Levels of Evidence Working Group. The Oxford 2011 levels of evidence. www.cebm.net/wp-content/uploads/2014/06/CEBM-levels-of-evidence-2.1.pdf (accessed 2017 Feb 21).
6. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med.* 2000; 342:1878-86.
7. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med.* 2000; 342:1887-92.
8. Concato J, Lawler EV, Lew RA et al. Observational methods in comparative effectiveness research. *Am J Med.* 2010; 123:e16-23.
9. Fleurence RL, Naci H, Jansen JP. The critical role of observational evidence in comparative effectiveness research. *Health Aff (Millwood).* 2010; 29:1826-33.
10. Marko NF, Weil RJ. The role of observational investigations in comparative effectiveness research. *Value Health.* 2010; 13:989-97.
11. Winterstein AG, Vermeulen LC. A tool kit for practice-based research. *Am J Health-Syst Pharm.* 2008; 65:1607.
12. Weber RJ, Cobaugh DJ. Developing and executing an effective research plan. *Am J Health-Syst Pharm.* 2008; 65:2058-65.
13. Smith KM. Building upon existing evidence to shape future research endeavors. *Am J Health-Syst Pharm.* 2008; 65:1767-74.
14. Planas LG. Intervention design, implementation, and evaluation. *Am J Health-Syst Pharm.* 2008; 65:1854-63.
15. Lipowski EE. Developing great research questions. *Am J Health-Syst Pharm.* 2008; 65:1667-70.
16. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health-Syst Pharm.* 2008; 65:2276-84.
17. Hartung DM, Touchette D. Overview of clinical research design. *Am J Health-Syst Pharm.* 2009; 66:398-408.
18. Gerhard T. Bias: considerations for research practice. *Am J Health-Syst Pharm.* 2008; 65:2159-68.
19. Devine EB. The art of obtaining grants. *Am J Health-Syst Pharm.* 2009; 66:580-7.
20. De Muth JE. Overview of biostatistics used in clinical research. *Am J Health-Syst Pharm.* 2009; 66:70-81.
21. De Muth JE. Preparing for the first meeting with a statistician. *Am J Health-Syst Pharm.* 2008; 65:2358-66.
22. Cobaugh DJ, Allison JJ. Understanding research principles: giving our patients the care they deserve. *Am J Health-Syst Pharm.* 2009; 66:1265.
23. Byerly WG. Working with the institutional review board. *Am J Health-Syst Pharm.* 2009; 66:176-84.
24. Motheral B, Brooks J, Clark MA et al. A checklist for retrospective database studies—report of the ISPOR Task Force on Retrospective Databases. *Value Health.* 2003; 6:90-7.
25. Farrugia P, Petrisor BA, Farrokhyar F, Bhandari M. Practical tips for surgical research: research questions, hypotheses and objectives. *Can J Surg.* 2010; 53:278-81.
26. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club.* 1995; 123:A12-3.
27. Reade MC, Eastwood GM, Bellomo R et al. Effect of dexmedetomidine added to standard care on ventilator-free time in patients with agitated delirium: a randomized clinical trial. *JAMA.* 2016; 315:1460-8.
28. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social,

behavioral, and biomedical sciences. *Behav Res Methods*. 2007; 39:175-91.
29. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999; 10:37-48.
30. Shahar E, Shahar DJ. Causal diagrams and the logic of matched case-control studies. *Clin Epidemiol*. 2012; 4:137-44.
31. Shahar E, Shahar DJ. Causal diagrams and the cross-sectional study. *Clin Epidemiol*. 2013; 5:57-65.
32. Sternbach GL. The Glasgow Coma Scale. *J Emerg Med*. 2000; 19:67-71.
33. Harris PA, Taylor R, Thielke R et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009; 42:377-81.
34. Kaji AH, Schriger D, Green S. Looking through the retrospectoscope: reducing bias in emergency medicine chart review studies. *Ann Emerg Med*. 2014; 64:292-8.
35. International Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals (updated December 2016). www.icmje.org/recommendations/ (accessed 2017 Feb 26).
36. Benchimol EI, Smeeth L, Guttmann A et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015; 12:e1001885.
37. Huber VC, Vogt HB. So you want to be an author: a primer on writing for publication in the medical literature part I: manuscript preparation. *S D Med*. 2016; 69:172-5.
38. Liumbruno GM, Velati C, Pasqualetti P, Franchini M. How to write a scientific manuscript for publication. *Blood Transfus*. 2013; 11:217-26.
39. O'Connor TR, Holmquist GP. Algorithm for writing a scientific manuscript. *Biochem Mol Biol Educ*. 2009; 37:344-8.
40. Sauaia A, Moore EE, Crebs JL et al. The anatomy of an article: title, abstract, and introduction. *J Trauma Acute Care Surg*. 2014; 76:1322-7.
41. Singer AJ, Hollander JE. How to write a manuscript. *J Emerg Med*. 2009; 36:89-93.
42. Welch HG. Preparing manuscripts for submission to medical journals: the paper trail. *Eff Clin Pract*. 1999; 2:131-7.

Appendix A—Checklist for retrospective data collection process

Item		Comments
☐	Pertinent variables available	• Construct causal diagram based on previous studies and clinical expertise. • Important variables must be reliably documented in the medical record.
☐	Systematic subject identification	• Define inclusion and exclusion criteria. • Minimize selection bias by chronological selection of subjects (if applicable). • Report number and reasons for patient exclusion.
☐	Standardized data collection instrument	• Construct a codebook for all variables. • Minimize free text (i.e., all variables should be categorical or numeric to enable analysis). • Test instrument on sample of charts, and modify as appropriate. • Use electronic data capture system such as REDCap to minimize errors.
☐	Training of data collectors	• Ensure that data collectors have adequate background to understand and interpret variables. • All abstractors should receive training (e.g., where variables are located in medical record) to ensure consistency and accuracy of collection. • Maintain communication with data abstractors during collection, and troubleshoot as needed.
☐	Ensure data accuracy and reliability	• All data should be double-checked and corrected by a second investigator via consensus. • Alternatively, 10% of sample can be checked by a second investigator for key variables (e.g., predictors, confounders, outcomes); test and report these variables for interrater agreement (e.g., kappa statistic, interclass correlation coefficient). • Check accuracy of outlier values (e.g., extremely low or high patient weight).
☐	Blinded collection of outcome variable	• Data collectors should not be privy to the study hypothesis (may not be possible). • Alternatively, collect outcome data separately using a different form and personnel; this could minimize potential bias in some studies where the treatment allocation is not readily apparent.

Appendix B—Authorship criteria and scoring system[a]

Criterion[b]	Task	Investigators						Maximum Points[c]
		A	B	C	D	E	F	
1. Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of the data	Study conception							10
	Study design							10
	Data acquisition							10
	Data analysis							10
	Data interpretation							10
2. Drafting the work or revising it critically for important intellectual content	Manuscript preparation							10
	Total score[d]							60
3. Final approval of the version to be published								
4. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved								

[a]Individuals must meet all 4 criteria to be considered for authorship; criteria 1 and 2 are scored (0–10 points per task).
[b]From reference 35.
[c]Maximum item point values may be adapted to give more weight to certain tasks.
[d]Total score determines authorship sequence (highest scorer is first author).

Appendix C—Checklist for manuscript preparation[a]

Section		Comments
Introduction	☐	• 2- or 3-paragraph structure • Paragraphs 1 and 2 (background, significance, rationale of study) discuss current evidence (main findings) and what the study intends to address • Paragraph 3 (primary and secondary objectives of study) states study hypotheses, if applicable
	☐	• Usual word range is 300–400 • The introduction is meant to be relatively brief and not a thorough review of the literature
Methods	☐	• Study design and setting • Type of study (e.g., retrospective cohort, case–control, cross-sectional). • Description of the setting to help readers assess external validity (e.g., "tertiary medical center in the United States"); the name of the institution should not be mentioned • Specific protocols, services, or facilities pertinent to the study (e.g., how the medications in the study are selected for patients) • Wording to indicate institutional review board approval
	☐	• Patient selection • Time period of the study (e.g., date range during which patient list was generated) • Method by which patient list was generated and screened and subjects were identified (e.g., by using diagnostic codes, medication lists) • Inclusion and exclusion criteria • Process for matching of cases and controls, if applicable

	☐	• Data collection processes
		• For retrospective studies, the process used for all elements mentioned in Appendix A (with the exception of patient selection) is explained and may include
		  ◦ Variables collected and how they were coded or defined
		  ◦ Training of data collectors
		  ◦ How data were collected
		  ◦ Methods used to ensure accuracy and reliability of data
		  ◦ Blinding techniques used for study outcome

Continued on next page

Continued from previous page

	☐	• Study objectives and outcomes • Primary, secondary, and other objectives and hypotheses • Definition used for primary and secondary outcome measures
	☐	Description of data analyses (ideally, this should be written with the help of the biostatistician involved in the project; components may include • Summary statistics (e.g., continuous variables were reported as means with S.D. values) • Statistical tests used for all analyses, including the primary and secondary outcomes • Regression methods used and variables considered to be pertinent confounders (include how confounders were selected) • Sample size determination and assumptions used • Software used for analyses
Results	☐	First paragraph indicates the number of patients included in the final sample and how they were identified: • List initial number evaluated and the numbers excluded for each of the listed exclusion criteria (alternatively, this can be accomplished by use of a flow diagram) • Summarize pertinent demographics or clinical parameters as necessary for the overall sample (i.e., give the reader a sense of the sample being investigated); may highlight imbalances between groups • Minimize overlap between data in tables and text
	☐	Main results: • Report main findings of the study • The results should flow in a manner that is sequential and consistent with the methods (i.e., report primary outcomes first, then secondary outcomes) • Ensure that the data analyses are described in the methods and also reflected and addressed in the results (and vice versa)
Discussion	☐	Paragraph 1: • Summarize the most important findings and implications of the study Paragraph 2 and beyond: • Discuss previous pertinent studies and comment on similarities and differences with the current investigation • Provide insight, interpretations, and applications of the results to practice Limitations: • Address issues regarding retrospective design (e.g., dependent on accuracy of documentation in the medical record) • Discuss any vulnerabilities that could affect the results (e.g., imbalances between groups, potential biases) Conclusions (may be separate section): • The conclusions should directly pertain to the primary and secondary objectives • Retrospective studies are hypothesis generating or exploratory; wording used should not overstate the conclusions

[a]Use this checklist in conjunction with STROBE and RECORD statements and ICMJE guidelines.