89 Clinical Research

KATE LESLIE, COR J. KALKMAN, and DUMINDA N. WIJEYSUNDERA

KEY POINTS

- Clinical research describes the characteristics and mechanisms of disease and injury in humans and investigates drugs, devices, diagnostic products, and bundles of care with the aim of providing high-quality evidence to guide clinical practice that improves patients' lives.
- Investigators are constrained by the need to draw conclusions based on a sample of patients, providers, or systems rather than the whole of the population of interest, and this introduces random, systematic, and design error that threatens internal and external validity.
- Observational (or nonexperimental) studies involve allowing nature or clinical care to take its course, without any major modification due to study-related procedures. However, nonexperimental designs are prone to systematic error, such as selection and information bias.
- In experimental studies the investigators do not let nature (or clinical care) take its course, but actively intervene to test a new intervention. Randomization and blinding reduce the risk of random and systematic errors in experimental studies, but generalizability can be limited.
- Systematic reviews synthesize the medical literature using a transparent search strategy that every reader can replicate and update, and accompanying metaanalyses produce aggregate estimates of treatment effects.
- A carefully designed protocol is the foundation of all clinical research, facilitating the review, conduct, and eventual publication of the project. Prior planning of sample size and statistical analyses is essential.
- Appropriate ethics, registration, and regulatory approvals; financial, data, and human resource management; plans for monitoring patient safety and data integrity; and plans for publishing the results and sharing the data are vital to the proper conduct of clinical research.
- A quality improvement cycle of reflection, feedback, and forward planning is beneficial. Giving patients an opportunity to comment on the studies in which they participated gives researchers fresh insight and opens up new avenues of research.

The measure of greatness in a scientific idea is the extent to which it stimulates thought and opens up new lines of research.
PAUL A.M. DIRAC

Introduction

Research is systematic investigation for the creation of generalizable knowledge. Clinical research describes the characteristics and mechanisms of disease and injury in humans, and investigates drugs, devices, diagnostic products, and bundles of care for human use. Clinical research also includes investigating the interactions of health professionals, students, and other stakeholders with the health care and health education systems. Most studies address an aspect of quality of care or education such as safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity.[1]

Miller's Anesthesia includes references to tens of thousands of clinical research studies. The main purpose of this chapter is to describe how these studies were designed and conducted. The chapter connects directly with Chapter 91, which elaborates on how to interpret and use evidence for clinical decisionmaking. Other related chapters include Chapter 1, Chapter 2, Chapter 4, Chapter 5, Chapter 8, and Chapter 30. Our goal is to describe how clinical research originates and to inspire our readers to create and support the use of high-quality evidence that is essential in guiding clinical practice that focuses on improving patients' lives.

Key Principles

Researchers want to produce results that will be useful to clinicians in improving patients' lives. However, they are constrained by the need to draw conclusions based on a sample of patients rather than the population of interest as a whole.[2] The random, systematic, and design errors introduced by studying a sample affect the confidence with which the association found in the study can be claimed to truly represent the causal effects of the exposure on the outcome ("internal validity") and impair the generalizability of the results to the source population for the sample and the population of interest as a whole ("external validity").[3,4] These errors will be discussed briefly here and will be highlighted in subsequent sections of this chapter.

RANDOM ERROR

Random error (or "chance") in the association between an exposure and an outcome is introduced by variation between the individuals in the study and within individuals

over time, and variation between and within measurements made during the study.[5,6] Random error is equally likely to cause false acceptance and false rejection of the null hypothesis (i.e., the hypothesis that there is no effect). The risk of drawing false conclusions increases with decreasing sample size and more statistical testing.

SYSTEMATIC ERROR

Systematic errors (or "biases") also threaten identification of the true causal effect that the exposure had on the outcome in the sample. They can arise from nature (confounding), selection features of the study (selection bias), or measurement features of the study (information bias). In contrast to random error, which distorts the results of a study in both directions, systematic error distorts only in one direction (i.e., toward acceptance *or* rejection of hypotheses).

Confounding

Confounding is a bias that arises when an association between an exposure and an outcome fails to take account of a third factor (a "confounder") that is associated with both the exposure and the outcome.[5,6] Confounders can be risk factors, preventive factors, or surrogate markers for another cause of the outcome, but they cannot be intermediary steps in the causal pathway between an exposure and an outcome. They may be known and measured ("known knowns"), known and unmeasured ("known unknowns"), or unknown and unmeasured ("unknown unknowns").[7] In observational studies, design (e.g., sampling restriction, matching) and analysis (e.g., statistical adjustment) options only address the distribution of measured confounders. Randomization distributes measured and unmeasured confounders evenly between groups (with increasing reliability in larger studies).[8] Concealment of allocation prevents selective recruitment, and blinding prevents confounding that results when knowledge of group allocation changes processes of care or behavior.[9]

The importance of confounding varies with the type of study. In studies focusing on causality (e.g., whether smoking causes lung cancer), investigators attempt to collect and manage all possible confounders, yet they must always take residual confounding into account or use randomization and blinding to ensure that known and unknown confounders are balanced between the groups. In studies focusing on prognosis (e.g., studies predicting the likelihood that a patient will survive the proposed surgical procedure) any variable that contributes to improved prediction can be used in the prediction model.[10]

Selection Bias

Bias can be introduced by the methods used to select a population of interest, to identify and sample a source of such patients, to recruit and retain those patients, and to disseminate the results.[4] Observational studies are prone to bias with respect to selection of sources of exposed and unexposed patients, or cases and controls. Patients and their treating team can introduce bias by participating or not participating based on the exposures and outcomes under study. Prospective cohort studies and randomized trials are prone to bias with respect to differential loss to follow-up.[2] Studies without preplanned and publicized protocols and statistical analysis plans are prone to bias with respect to selection of outcomes to report (favoring outcomes with statistically significant differences between groups), and studies that fail to find a difference between groups with respect to the primary outcome ("negative" studies) may result in "publication bias," that is, they are either never submitted or not selected for publication by journal editors.[11]

Information Bias

This type of bias can be introduced by inaccurate measurement or classification of exposures, outcomes, and other measured variables. This error arises when survey instruments and diagnostic tests are invalid or unreliable and may be differential (affecting groups differently) or nondifferential (affecting groups similarly).[2] Recall bias arises when recollection of past exposures is affected by whether the patient experienced the outcome or not. Socially desirable response bias arises when participants give answers based on their assumptions about the investigators and society. Patients who are not assessed for the primary outcome (i.e., whose information is "missing") can cause random or systematic error, and this should be assessed with sensitivity analyses.[12]

DESIGN ERROR

Design errors limit the usefulness of research to clinicians, even if the research is free of random and systematic errors, by affecting its generalizability.[3] Examples of such design errors include studying exposures that are expensive or difficult to implement; comparing new treatments with placebos or weak treatments rather than the best available option; assessing outcomes that are not relevant to patients and the community; and seeking to prove that a new treatment is superior to an old treatment, when proving that it is equivalent or non-inferior would be more useful.[3]

STATISTICAL INFERENCE

The *P*-value is the "probability under a specified statistical model that a statistical summary of the data (e.g., the sample mean difference between two compared groups) would be equal to or more extreme than its observed value."[13] For example, $P = .05$ means that under the null hypothesis there was a 5% chance of observing results at least as extreme as those seen in the study.[14] *P*-values can indicate how incompatible the data are with a specified statistical model, but do not measure the probability that the studied hypothesis is true nor that the data were produced by random error alone.[13] Confidence intervals are better suited to reflecting the size and precision of the treatment effect than *P*-values: the 95% confidence interval refers to the fact that if the same study were repeated many times and the confidence intervals were similarly calculated for each case, 95% of such intervals would include the true treatment effect.[14] The minimal clinically important difference is important when assessing confidence intervals. If the lower limit of the confidence interval excludes the minimal clinically important difference, then the effect of the treatment is likely to be important.[14] Bayesian inference overcomes some of the limitations of *P*-values and confidence intervals.[15] Instead of interpreting the frequency of a phenomenon, Bayesian

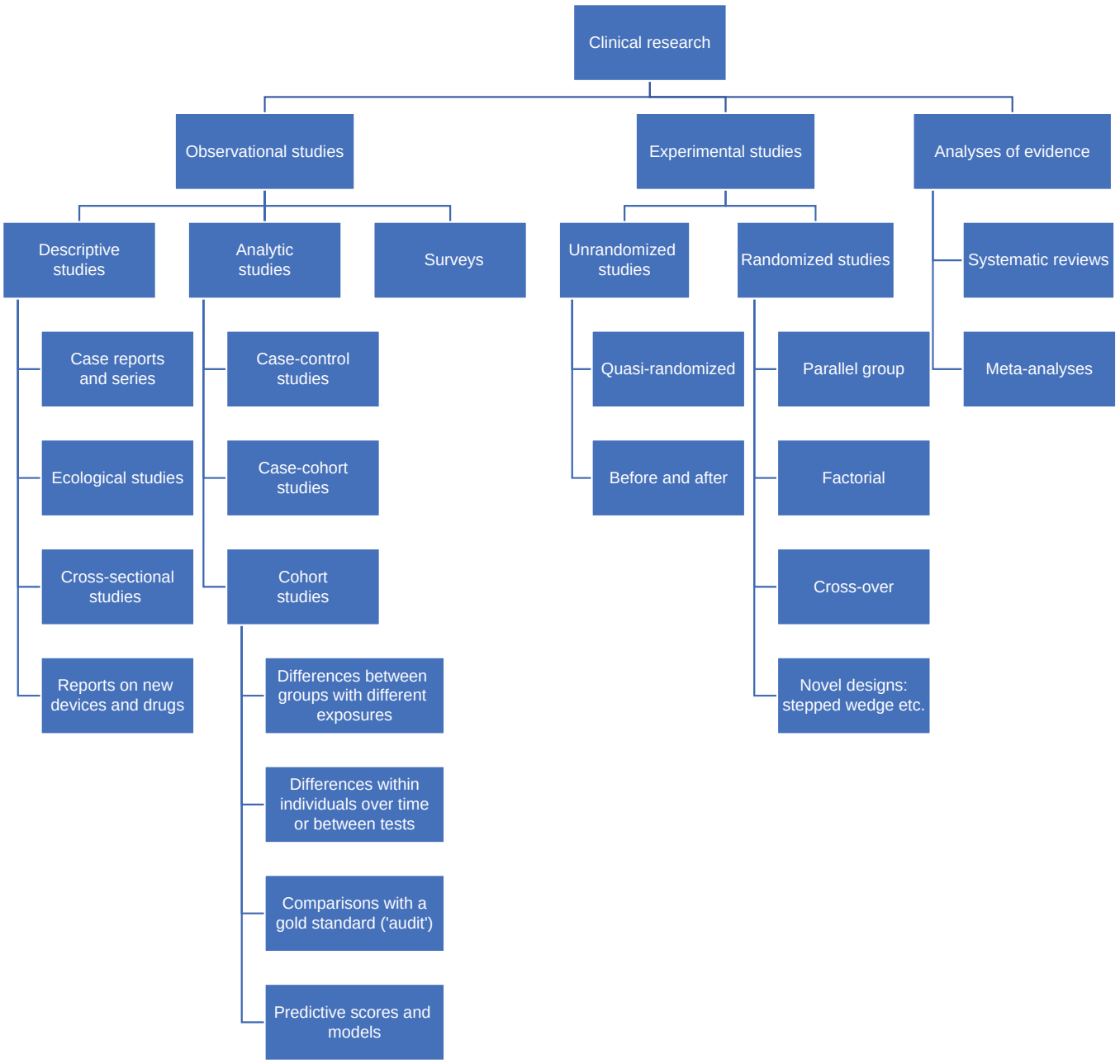

Fig. 89.1 Study designs in clinical research.

interference incorporates prior evidence, biological plausibility, and pre-existing beliefs into the calculation of the probability of a treatment effect.[14]

Study Design

The study designs used in clinical research are outlined in Fig. 89.1 and Box 89.1. In this chapter, we use the terms "retrospective" and "prospective" to describe the timing of measurement of the exposure and outcome relative to the start of the study, rather than to describe the direction of enquiry (i.e., outcome → exposure or exposure → outcome). We use the US National Institutes of Health definition of a clinical trial ("any research study that prospectively assigns human participants or groups of human participants to one or more health-related interventions to evaluate the effects on health outcomes").[16]

OBSERVATIONAL STUDIES

Observational (or nonexperimental) studies involve allowing nature or clinical care to take its course, with assignment to the intervention based on usual practice, and without any major modification due to study-related procedures, such as recruitment or data collection.[17] An observational study is not a clinical trial, therefore the contradictory term "observational trial" should be avoided. Ethical and pragmatic considerations often mean that an observational design is the only approach to answer a research question.

BOX 89.1 Study Design in Clinical Research: Depth of Anesthesia and Awareness

A ***case study*** reported on a patient who received light anesthesia and was aware.
An ***ecological study*** reported on the incidence of awareness in hospitals with low or high per-patient volatile anesthetic use.
A ***survey*** asked anesthesiologists about their routine use of light or deep anesthesia and the incidence of awareness in their practice.
A ***case-control*** study found patients with awareness, matched them with patients without awareness, and then determined if they had received light or deep anesthesia.
A ***case-cohort*** study found patients who had light anesthesia, matched them with patients who had deep anesthesia, then determined if they had been aware.
A ***retrospective cohort*** study examined existing records to determine the incidence of awareness in patients who had received light or deep anesthesia.
A ***prospective cohort*** study enrolled patients having general anesthesia and followed them to determine the incidence of awareness in patients receiving light and deep anesthesia.
A ***within-patient comparison*** of two processed EEG monitors determined each monitor's ability to predict awareness.
An ***audit*** compared current practice with a national evidence-based guideline on prevention of awareness during general anesthesia.
A ***prediction study*** reported on the development of a risk score for awareness based on a large cohort of patients having light and deep anesthesia.
A ***before-and-after*** study enrolled a cohort of patients having anesthesia without EEG monitoring and then a cohort having anesthesia with EEG monitoring and compared the incidences of awareness.
A ***randomized controlled trial*** randomized patients to light anesthesia or deep anesthesia and compared the incidence of awareness in each group.
A ***systematic review*** searched for clinical trials examining the relationship between light and deep anesthesia and awareness, and conducted ***metaanalysis*** to determine the pooled risk of awareness in patients receiving light or deep anesthesia.

EEG, Electroencephalographic.

For example, most research evaluating the impact of cigarette smoking on lung cancer relies on nonexperimental methods,[18] since assigning smoking as an intervention is unethical.[18] In general, observational studies can demonstrate associations between exposures and outcomes, but do not necessarily prove a causal relationship. Several factors must be carefully considered, including the consistency of findings across multiple high-quality studies, biological plausibility, and whether clear dose-response and temporal relationships exist between exposures and outcomes.[19] This is the case for cigarette smoking: the weight of observational research eventually established cigarette smoking as a cause of lung cancer.[18]

Since most biomedical research is observational in design, several initiatives have been undertaken to improve the conduct and reporting of such studies. For example, the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement makes recommendations for the reporting of three specific types of observational studies, namely cross-sectional, case-control, and cohort studies.[20] However, strong evidence that these tools improve the design and reporting of observational studies is lacking.[21]

Descriptive Studies

Descriptive studies aim to describe the characteristics of patients in a sample, such as baseline features (e.g., age, sex, comorbid disease), processes of care (e.g., anesthesia type), and outcomes (e.g., mortality, stroke). Unlike analytic studies, they do not seek to define relationships among these characteristics.

Case Reports and Series. Case reports describe exposures and outcomes in individual patients, while case series provide the same information for groups of patients. Case reports and series are suited to describing newly recognized uncommon exposures and outcomes, such as drug-related adverse effects.[22] As these studies focus on individuals with a specific exposure or outcome, no direct comparisons are possible with unexposed individuals or those without the outcome. Despite these inherent limitations, case reports and series have made important contributions to medicine. For example, a case report first identified lipid emulsion as a successful treatment for local anesthetic cardiotoxicity[23] and a case series of anesthetic-related deaths in a family first identified the pharmacogenetic disorder malignant hyperthermia.[24]

Descriptions of New Interventions. These studies are a subtype of case reports and series that describe new interventions (e.g., drugs, devices, diagnostic products, monitors, bundles of care, survey instruments, guidelines), but without any comparison against a control group. Such studies have provided valuable information to advance medicine. Important examples include the first use of the non-depolarizing muscle relaxant curare[25] and the initial clinical experience with the laryngeal mask airway.[26] With respect to diagnostic tests, "test research" determines the properties of the test itself (i.e., sensitivity, specificity, likelihood ratios, positive and negative predictive values). However, predictive values in particular are dependent on the prevalence of the outcome in the study population. A better approach ("diagnostic research") determines the extent to which adding a new diagnostic test improves the likelihood of arriving at the correct diagnosis compared with existing diagnostic criteria.[27] Diagnostic tests should always be evaluated using both methods, especially if the new test is simpler, cheaper, less invasive, or less burdensome to the patient than the existing test.[28]

Analytic Studies

Analytic studies seek to measure associations between exposures and outcomes. In contrast to descriptive studies, where all patients have the specified exposure (e.g., a laryngeal mask airway[26]) and/or outcome (e.g., successful resuscitation from local anesthetic cardiotoxicity[23]), in analytic studies patients with and without the exposure of interest (e.g., early childhood exposure to anesthesia) and outcome of interest (e.g., poor educational performance) are required in order to assess associations beyond those expected by chance alone.[29]

Ecological Studies. Ecological studies or "aggregate risk studies" measure exposure status and outcome status as average values across groups of individuals. They are particularly suited to exposures and outcomes that are routinely measured in populations. A major limitation of this approach is that exposure status and outcome status may not be linked at the individual level, leading to the "ecological fallacy," where individuals within the group-level unit with the outcome of interest may not have been the individuals with the exposure of interest. The potential for this bias was highlighted in a study of hospital-level use of neuraxial anesthesia for hip fracture surgery.[30] Higher hospital-level rates of neuraxial anesthesia use were associated with lower overall mortality, but there was no evidence of such an association at the individual level. In fact, at hospitals with higher rates of neuraxial anesthesia use, all patients had better outcomes after hip fracture surgery, regardless of the type of anesthesia they received, suggesting that the causal mechanisms for hospital-level effects were factors other than neuraxial anesthesia use.

Cross-Sectional Studies. Cross-sectional studies assess the exposure status and outcome status of individuals at the same time (or within a short and stable interval). They are suited to exposures that do not change over time, such as genetically-determined characteristics or chronic stable health conditions. Cross-sectional methods have been used to determine whether tibial nerve ultrasound can detect diabetic peripheral neuropathy[31] and whether a preoperative screening questionnaire detects obstructive sleep apnea.[32] Most surveys (see later) are cross-sectional studies since respondents typically complete the survey at a single time-point. Cross-sectional studies are ill-suited to establishing causal relationships because of the "causality dilemma" (i.e., they do not permit clear delineation of the temporal link between putative exposures and outcomes). For example, children presenting for dental treatment and their accompanying parent were tested for anxiety. Forty percent of the children and 60% of the parents were anxious preoperatively, but the investigators could not determine with any certainly if these states were related and, if so, the direction of effect.[33]

Case-Control Studies. Case-control studies assemble participants based on their outcome status; that is, patients who experienced the outcome (cases) and patients who did not (controls). Once the sample is assembled, exposure status is ascertained by looking back in time (Fig. 89.2*A*). Case-control studies are inexpensive, amenable to quick completion, and suited to the study of rare outcomes, such as postoperative stroke[34] and ischemic optic neuropathy.[35] However, they are susceptible to selection bias, especially in relation to the controls, who should be drawn from the same underlying population as the cases.[36] Thus, in a hypothetical study of stroke following cardiac surgery, it would not be appropriate to select cases who had complex aortic arch procedures and controls who had straightforward coronary artery bypass grafting procedures. Investigators can further avoid selection bias by matching cases with up to five controls based on prognostically important characteristics (e.g., age, sex), using more than one control group, and measuring exposure status with equal rigor across cases and controls (avoiding information bias).

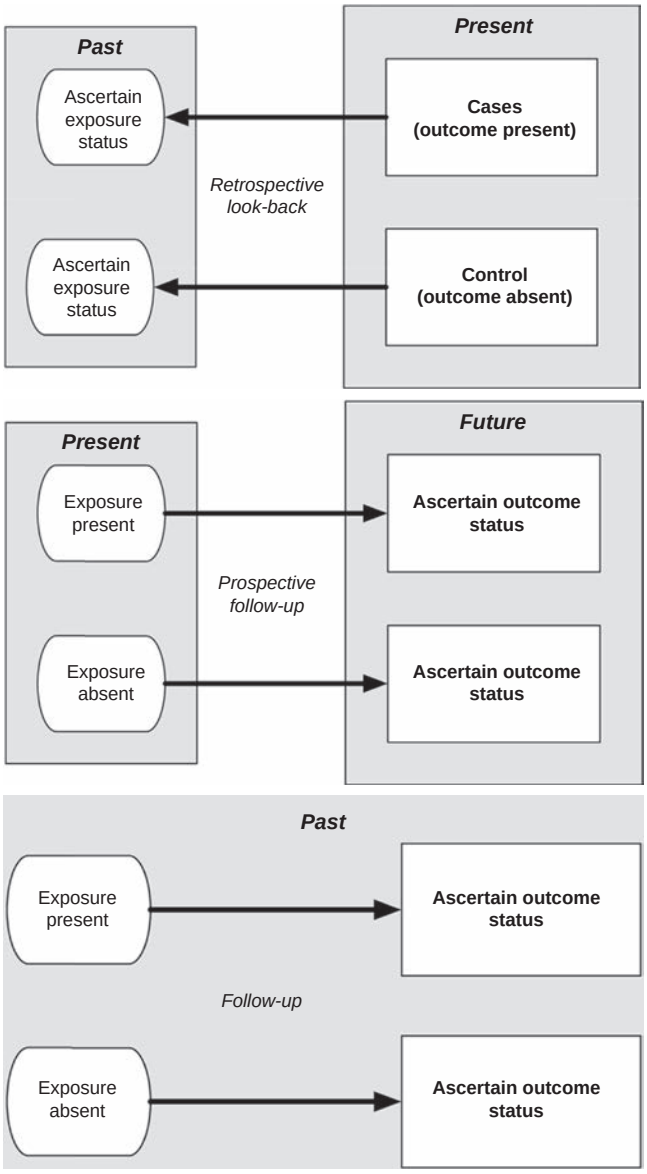

Fig. 89.2 Observational designs.

Cohort Studies. Cohort studies assemble participants based on their exposure status and follow them forward in time to ascertain the presence of outcomes. Both the exposure and outcome need to be relatively common to limit the size of the cohort required. Cohort studies can be conducted prospectively and retrospectively. In a prospective study, the cohort is assembled in the present and followed into the future (see Fig. 89.2*B*). This allows investigators to carefully measure exposure status (without recall bias based on knowing the outcome); delineate temporal relationships between exposures and outcomes; and implement standardized follow-up. Standardized follow-up is particularly important when the outcomes might be missed or when baseline characteristics might influence surveillance for the outcomes. For example, standardized troponin estimation is required to detect myocardial infarction and injury, because these outcomes may be silent and some patients are less likely to be tested than others (i.e., females, low-risk surgery patients).[37] Prospective cohort studies are more

expensive and take longer to complete than case-control studies and retrospective cohort studies.

In retrospective cohort studies, cohort assembly and follow-up occur entirely in the past (see Fig. 89.2*C*). Very large databases make addressing uncommon exposures and outcomes feasible using this design. The data needed for the research (i.e., patient characteristics, processes of care and outcomes) can be captured from one or more pre-existing sources, including paper and electronic medical records, administrative and legal databases, government and clinical registries, and databases assembled for research purposes. In anesthesiology, increasing access to these sources has facilitated growth in such research.[38-40] Retrospective cohort studies are quicker and less expensive to conduct than prospective cohort studies. However, they are limited by the available data sources, which may vary in completeness and accuracy, and be influenced by the intensity of follow-up protocols. For example, the reported incidence of postoperative myocardial infarction varied considerably between prospective studies that used standardized surveillance and retrospective studies that used registries of usual clinical practice.[37,39]

Case-Cohort Studies. Case-cohort studies are a subtype of cohort studies where exposed patients are matched to unexposed patients, then followed forward in time to measure outcomes. They can be conducted either prospectively or retrospectively and are suited to the study of rare exposures. Matching reduces the influence of differences in prognostically important characteristics (e.g., age, comorbid status) between exposed and unexposed individuals. For example, a retrospective cohort study evaluating the impact of early childhood anesthesia on educational outcomes matched exposed and unexposed children based on gestational age at birth, maternal age at birth, year of birth, sex, and location of residence.[29] Matching on large numbers of baseline characteristics generally is not feasible because of the difficulties in identifying suitable individuals. In these cases, propensity score matching is an alternative that can help assemble a matched cohort with very similar baseline characteristics for both exposed and unexposed individuals.[41] Importantly, however, propensity score matching does not remove bias related to imbalances in unmeasured confounders between the groups.[41]

Studies Assessing Changes or Differences in Individuals. Some studies make serial measurements in individuals to assess changes over time. Examples include studies that characterize the pharmacokinetics and dynamics of anesthetic drugs[42,43] or postoperative acute pain trajectories.[44] These studies are a subtype of cohort study that incorporate longitudinal repeated measurement of an outcome measure. Statistical analyses of these data must account for correlated measurements in individuals over time.[45] Other studies make parallel measurements in individuals to assess differences between these measurements. Examples include comparing coagulation testing methodologies during surgery[46] or disability scoring instruments after surgery.[47] Bland-Altman analyses which test investigator-defined limits of agreement are an appropriate analysis technique when comparing tests,[48] while scales are often compared using correlation on measures such as validity, reliability, and responsiveness.[49] Some of these studies could equally be described as "experimental" (see later) because the investigators control the intervention.

Studies Assessing Practice Against a Gold Standard ("Audit"). Audits are variants of cohort studies that involve assembling a cohort of patients and determining whether practice complies with an external standard. The term "audit" is sometimes used inaccurately to describe studies that determine the standard that clinical practice achieves (i.e., evaluation) or compare practice with an investigator-derived standard (i.e., research). The extent of compliance with the standard can be compared based on different exposures. Examples of audits include the compliance of venous thromboembolism[50] and surgical site infection[51] prophylaxis with national guidelines. An important issue for such studies is ensuring that the external standard is reasonably valid and accepted within the wider community.

Studies Developing and Validating Prediction Tools. The goal of these studies is to develop tools (e.g., scores, prediction models, risk calculators) that accurately predict outcomes in individual patients. A good clinical prediction tool should be simple to use, exhibit good discrimination (i.e., the extent to which the tool correctly assigns higher predicted risk estimates to individuals who experience the outcome), and demonstrate acceptable calibration (i.e., the extent to which the observed outcome event rates agree with event rates predicted by the tool).[52] These studies first use statistical methods to evaluate associations between exposures (e.g., baseline characteristics, processes of care) and outcomes in one cohort of patients; use the results to develop scoring systems or prediction models; and then validate them in another cohort of patients.[52] Cohort study datasets are the preferred sources of development cohorts because randomized controlled trial datasets may not be generalizable.[53] Different statistical methods can be employed to develop these scores and models, including logistic regression modeling for categorical outcomes,[54] Cox proportional hazards modeling for time to event outcomes,[55] recursive partitioning (i.e., decision tree analysis),[56] and machine learning techniques (e.g., artificial neural networks).[52] Examples from the perioperative setting include the Assess Respiratory Risk in Surgical Patients in Catalonia (ARISCAT) score for predicting postoperative pulmonary complications,[57] the American College of Surgeons National Surgical Quality Improvement Program (NSQIP) calculator for predicting surgical risk,[58] and the Acute Physiology and Chronic Health Evaluation (APACHE) score for predicting hospital mortality in critically ill patients.[59]

Surveys

Surveys collect information from individuals (patients, families, staff, students) or organizations (hospitals, universities, employers) about facts and attitudes.[60] Most are cross-sectional studies assessing exposures and outcomes at the same time. Factual surveys ask for information or test knowledge. Attitudinal surveys ask about attitudes, beliefs, and intentions. Surveys can be descriptive (describing responses from the whole group) or analytic (comparing

responses in sub-groups). Surveys must be carefully planned and executed in order to protect participants and provide reliable conclusions.[60,61] Response bias (non-response, incorrect response, or untruthful response) is a particular issue with survey research and must be explicitly considered in survey design and reporting.[61-63] A systematic review of 240 surveys published in six anesthesia journals revealed that reporting was inconsistent, particularly with respect to articulating a hypothesis, describing the design, calculating sample size, providing confidence intervals, and accounting for non-responders.[64] Compliance with design and reporting checklists may improve the quality of survey research.[20,61,62]

Health Services Research

Health services research (also known as health systems research or health policy and systems research) has been defined as a "multi-disciplinary field of scientific investigation that studies how social factors, financing systems, organizational structures and processes, health technologies, and personal behaviors affect access to health care, the quality and cost of health care, and ultimately our health and well-being. Its research domains are individuals, families, organizations, institutions, communities, and populations."[65] While typical clinical research focuses on the epidemiology, risk factors, prognosis, and interventions related to specific disease entities (e.g., ischemic heart disease, postoperative respiratory failure), health services research focuses on the delivery of health care. Health services research and clinical research differ with respect to their focus, but they have considerable overlap with respect to their research methodologies (i.e., how studies are designed to answer specific research questions). Thus, health services researchers also employ surveys, observational designs (as described previously), and experimental designs (as described in the section to follow). In addition, qualitative research methods are employed, such as thematic analyses of individual interviews and focus groups. Qualitative methods are particularly suited to identifying potential underlying reasons for clinician and patient behaviors within healthcare settings. Examples of health services research in the perioperative setting include a retrospective cohort study to evaluate variation in rates of preoperative medical consultation for major surgery,[66] a prospective cohort study of critical care utilization after major surgery,[67] a mixed qualitative-quantitative methods study of a standardized operating room to intensive care unit handoff process,[68] and a stepped wedge cluster randomized trial of a multifaceted implementation of perioperative safety guidelines.[69]

EXPERIMENTAL STUDIES

In experimental studies, the investigators do not let nature (or clinical care) take its course, but actively intervene to test a new intervention. Experimental studies are almost always of a parallel group design where patients or clusters of patients are allocated to an intervention or control treatment (i.e., placebo, usual care, or current best practice) by the investigators and then tested for outcomes (Table 89.1). Newer designs include cluster randomized, factorial, stepped wedge, and adaptive studies.

TABLE 89.1 Advantages and Disadvantages of Controlled Study Designs

	ALLOCATION	
Timing of Data Collection	**Randomized**	**Nonrandomized**
After intervention	Any baseline differences could bias the results.	Comparisons may be confounded by differences between departments or organizations at baseline.
Before and after intervention	Allows for specific comparison of change net of any baseline differences. Enables comparisons to be made between sites that change most or least.	Controls for baseline differences possible. Rates of change are less confounded than cross-sectional data.

Modified from Brown C, Lilford R. Evaluating service delivery interventions to enhance patient safety. *BMJ*. 2008;337:a2764.

Unrandomized Studies

If the investigator allocates the intervention or control to study patients in a nonrandom manner, this introduces selection bias and potential for imbalance between the groups at baseline in terms of risk for the primary outcome.[70] Overestimation or underestimation of the true treatment effect may then occur.

Quasi-Randomized Studies. Quasi-randomized (or quasi-experimental) studies attempt to select patients for the intervention or control in a less obvious but still nonrandom manner, for example, by surgical specialty, day of the week, date of birth, or by using a cut-off score for a certain characteristic. Quasi-randomized designs are seldom considered acceptable in contemporary clinical research because it is very hard to conceal the allocation, prevent selection bias, and ensure blinding. Some quasi-randomized designs may allow limited inference about causality (e.g., those that include a pre-test in the intervention and control groups to establish comparability). Sometimes randomization is not possible and quasi-randomized designs may be the only way to address the research question.[70] The Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I) tool helps readers to assess the quality of observational studies that evaluate interventions without the benefit of randomization.[71]

The before-and-after study is a simple and tempting study design that has been used for centuries to compare patient status before and after a new intervention. For example, John Snow compared the incidences of cholera in a neighborhood of London before and after removing the local water pump handle.[72] In modern research, these studies can be conducted entirely prospectively, partly prospectively (the "after" group), partly retrospectively (the "before" group), or wholly retrospectively.[73] The "before" intervention is often current usual care and the "after" intervention is often a new treatment. For example, an enhanced recovery pathway for colectomy patients was reported to be associated with decreased opioid administration, earlier return of bowel function, and reduced length of hospital stay compared to the old

approach.[74] The methodological challenge here is the influence of time: these improvements could have resulted from the new pathway or from changes in case-mix or unmeasured parallel improvements in care. If the effects of such parallel improvements are large enough, they might even obscure the fact that the new pathway is actually worse in terms of patient outcomes. Difference-in-differences approaches can be used to address time-related trends in outcomes. These approaches assume that trends unrelated to the exposure are the same in both groups.[75] Before-and-after studies are particularly vulnerable to the Hawthorne effect (i.e., the observation that people perform better when they know that they are being observed). The observational study design in which data are collected in individual patients before and after an intervention is addressed above.

Cross-Over Studies. The cross-over study is a more robust form of the before-and-after design. Here each patient receives all the interventions and control treatments, separated by wash-out periods to eliminate carry-over effects from the previous treatment.[76] Cross-over studies are well suited to patients taking chronic medications (e.g., analgesics for chronic pain[77]). A major benefit of the cross-over design is that patients are their own controls, which eliminates the confounding issues seen in parallel group designs. The order in which patients receive the intervention and control treatments can be nonrandom or random (the latter will eliminate time effects). In addition, administration of the intervention and control treatments can be blinded to eliminate placebo and nocebo effects.[78]

Randomized Studies

The ascendency of large simple randomized trials as the most robust form of primary research in anesthesia, intensive care, and pain medicine had its roots in the evidence-based medicine movement.[79] The cardinal features of a high-quality randomized controlled trial are successful randomization and blinding, and a sufficient sample size to reveal clinically important treatment effects on patient-centered outcomes.

Randomization. The many methodological problems associated with observational studies and nonrandomized experimental studies boil down to the distorting influence of confounders (see above).[80] The purpose of random allocation is to ensure that all confounders (known and unknown) are evenly distributed between the intervention and control groups at baseline. The success of randomization in evenly distributing these characteristics is critically dependent on the sample size of the study. After successful randomization any remaining differences in baseline characteristics will be the result of chance alone and patients in the intervention and control groups should have the same probability of experiencing the primary outcome. The study will then have a small likelihood of distorted results from random and systematic error. This is the primary reason that large randomized trials are considered the gold standard for experimental studies in medicine.[79] Randomization schedules can be generated using simple, block, stratified, or covariate adjusted techniques.[81] Allocation concealment is an essential step in preventing selection bias (i.e., recruiting patients based on knowledge of the treatment to which the next study patient will be allocated).[82]

BOX 89.2 Phases of Clinical Trials

Phase I

Phase I trials test a new intervention for the first time in a small group of participants to evaluate safety.

Phase II

Phase II clinical trials study a new intervention in a larger group of participants to determine efficacy and further evaluate safety.

Phase III

Phase III trials study the efficacy of a new intervention in a large group of people by comparing it to other interventions or routine care and monitor adverse effects.

Phase IV

Phase IV trials study an intervention after it has been marketed, and monitor its effectiveness in the general population, collect information about adverse effects, and investigate its use in a different condition or in combination with other therapies.

Blinding. Randomization can effectively eliminate confounding, but it is not enough to guarantee unbiased results. Simply knowing about the treatment allocation can influence the behavior of the investigators, subsequent clinical management by the treating team, and even the symptoms that the patients experience, creating a new imbalance in confounders and undoing the benefits of randomization. The solution here is blinding: hiding the treatment allocation from the observers collecting study data, the patient, and/or the treating team.[83] Blinding observers, patients, and treating team members in phase III and IV drug trials (Box 89.2) is the norm, as matched blinded experimental and control drugs are easy (although not inexpensive) to produce.[84] Other interventions are harder to blind (e.g., intravenous albumin solutions[85]) and others cannot be blinded (e.g., epidural local anesthetics in chronic pain patients[86]). In such cases, it is important to conceal allocation as long as possible to prevent this knowledge from influencing processes of care. Every effort should be made to blind outcome observers in randomized trials.

Superiority, Equivalence, and Non-Inferiority. Randomized trials are often geared toward demonstrating that the intervention results in better outcomes than the control ("superiority").[87] However, a randomized trial can be designed to demonstrate that the two treatments produce the same results ("equivalent")[88] or that the one produces results at least as good as the other ("non-inferiority").[89] Equivalence and noninferiority designs are useful if the new intervention might be simpler, safer, and/or cheaper than the current treatment, and demonstrating equivalence or non-inferiority would be sufficient to change practice.

Factorial Designs. Factorial designs allow testing of more than one intervention in a single clinical trial.[90] Rather than performing a separate randomized trial for each intervention and each combination, patients in factorial studies are separately randomized to two or more different interventions (that is, they receive none, some, or all of the experimental interventions). This design is efficient and allows testing of interactions between the interventions.

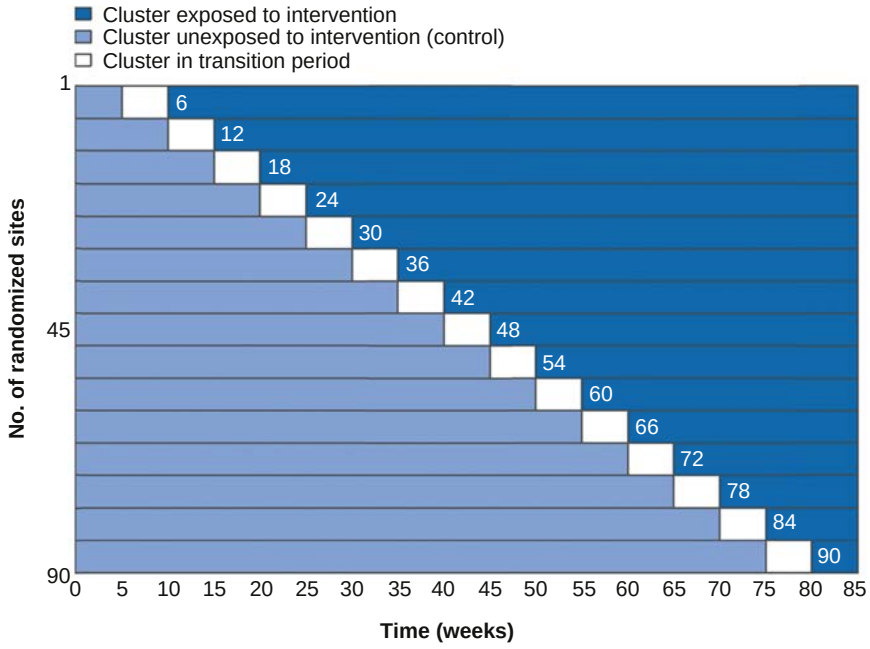

Fig. 89.3 Stepped wedge design.

For example, in a factorial trial of six interventions for the prevention of postoperative nausea and vomiting, 4000 patients were randomized to one of 64 possible combinations of six antiemetics.[91] This design allowed the investigators to conclude that these antiemetics were similarly effective and acted independently. In another example, 5784 patients were randomized to aspirin or placebo, and in a partial factorial, 4662 of these patients were also randomized to tranexamic acid or placebo.[92,93] This allowed the investigators to determine that there were no interactions between the interventions with respect to death, thrombotic complications, or major hemorrhage.

Cluster Randomized Designs. Most studies are randomized at the patient level. Cluster randomized trial designs are needed when randomization at the patient level is not possible or is methodologically unsound. This is particularly the case for process of care interventions because the fidelity of the intervention is dependent on the execution by the treating team and blinding is often not possible.[94] Examples include cross-over cluster randomized trials of selective decontamination of the gut in intensive care patients (even though the intervention is only applied in patients randomized to the treatment group, changes to bacterial colonization profiles will extend to all the patients in the unit)[95] and randomized trials of the introduction of medical emergency teams to treat deteriorating hospital patients.[96]

Stepped Wedge Designs. Cluster randomization means that some hospitals or clinical areas implement the new intervention, and some must remain with the existing model of care. Before-and-after studies are contaminated by the effects of time and cross-over cluster randomized trials have the disadvantage that clusters that were first randomized to the new intervention must revert to the existing model of care. A stepped wedge design circumvents some of these ethical and methodological issues by ensuring that each cluster first provides data for the control condition and then crosses over to the new intervention (Fig. 89.3).[97,98] This may improve recruitment of centers and patients. The duration of these periods differs for every cluster, but at the end of the study period there will be an equal amount of data from control and intervention periods. This eliminates—to a large extent—the effects of time. Stepped wedge designs were originally used in vaccination studies, exploiting the natural limitation that vaccination programs can never be rolled out over an entire region in a very short time frame.[99] They are considered alternatives for randomized trials of complex interventions, provided that there is a high likelihood of a positive effect of the intervention and a very low risk of harm.[97]

ANALYSES OF PUBLISHED RESEARCH

Systematic Reviews

The explosion in published medical knowledge and the requirement for quick answers to urgent clinical questions means that reliable synthesis of the medical literature is invaluable. Traditional narrative reviews are prone to author biases (sometimes even outright conflicts of interest), because authors can "cherry pick" the literature for evidence supporting their own opinions. Systematic reviews pose an explicit research question and publish a transparent search strategy that every reader can replicate and update.[100] The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement is available to guide authors of systematic reviews.[101] However, the quality of reporting and methodological quality of reviews remains inconsistent.[102]

Metaanalyses

When multiple randomized trials have addressed similar research questions, interventions, and outcomes, the results can be mathematically aggregated in a metaanalysis,[100] the goal of which is to estimate the aggregated effect of the intervention. A high-quality metaanalysis of large

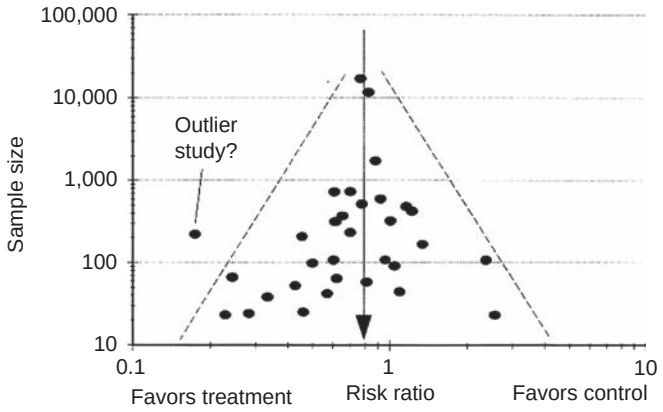

Fig. 89.4 Funnel plot showing publication bias: overrepresentation of studies with small sample sizes showing larger treatment effects.

well-conducted randomized trials is considered the highest level of evidence to guide practice, although results of metaanalyses do not always agree with subsequent very large trials.[103] The Cochrane Collaboration is dedicated to performing systematic reviews and metaanalyses that inform clinical practice. The Collaboration's software tools ("RevMan" and "Covidence") generate analyses, graphs, and metrics about the quality of included studies.[104,105] Funnel plots of treatment effect against study precision are used to detect publication bias toward studies with large treatment effects (typical of small trials) (Fig. 89.4).

Individual Patient Metaanalysis

A more powerful form of metaanalysis aggregates results using individual patient data, rather than aggregated data from the included studies.[106,107] The benefits of this approach are better characterization of outcomes and the opportunity to perform new subgroup analyses. The challenges of this approach include obtaining the raw data (with one paper reporting that only 25% of these metaanalyses retrieved all individual patient data[108]) and reliably merging trial databases containing the original data into the new pooled database. A standard aggregated data metaanalysis will illuminate whether an individual patient data metaanalysis will yield sufficient new information to justify the additional time and effort.

Study Protocol

A carefully designed protocol is the foundation of all clinical research, facilitating the review, conduct, and eventual publication of the study (Fig. 89.5).[109] However protocols are often incomplete and inconsistent with their corresponding published reports.[110,111] In response, templates have been developed, including by the World Health Organisation,[112] the International Council for Harmonisation,[113] national research funding agencies,[114] and the Equator Network (the Standard Protocol Items: Recommendations for Interventional Trials [SPIRIT] statement[115,116]). The effectiveness of these templates in improving the quality of protocols is yet to be determined. Amendments to clinical research protocols must be approved by the coordinating institutional review board (human research ethics committee) and must be reflected in the trial's registration. The International Committee of Medical Journal Editors recommends that investigators make their protocols publicly available before any results are published, for example by posting them on institutional websites or publishing them in peer-reviewed journals.[117]

HYPOTHESES

All clinical research—from a case report through qualitative and observational projects to a multicenter clinical trial—begins with a research question. These questions arise from a myriad of sources: the existing literature, the previous work of the investigators, interactions with colleagues, observations during clinical work, and discussions with patients and their families. A well formulated research question guides the literature review (which establishes the need for the study); informs the study design, methods, and sample size; and limits the potential for error and bias.[118] The Population-Intervention-Comparator-Outcome-Timeframe-Setting (PICOTS) format is a popular framework among clinical researchers[119] and improves the quality of research questions.[118] In an observational study, the "intervention" and "comparator" can be patient-based (e.g., with or without intraoperative hypotension) or care-based (e.g., treated in a metropolitan or rural hospital). In a clinical trial, the intervention and comparator are randomized. During protocol development, research questions are transformed into formal hypotheses. The structure of the hypothesis depends on whether the investigators predict that one group will have superior, equivalent, or non-inferior outcomes to the others. Regardless of this prediction, two-sided hypothesis testing is preferable: that is, entertaining the possibility that standard treatment is better or that the treatments are not equivalent.[120]

POPULATION

Selection of participants depends on the purpose and context of the study, and is a key differentiator between explanatory and pragmatic studies.[121,122] Explanatory (efficacy) studies, such as observational studies of new drugs and devices or phase I to III clinical trials, include highly selected patients in highly controlled settings, to reduce variability between patients and maximize treatment effects. The generalizability of explanatory studies therefore is limited. Pragmatic (effectiveness) studies, such as phase IV clinical trials, include typical patients receiving real-world care.[123] Random sampling is approximated by approaching all patients who have indications for the treatment and who present themselves for care. In practice, however, only a subset of patients is approached. Further nonrandom selection occurs when patients decline to participate or fail to complete data collection.[122] Although national funders promote diversity among participants,[124] children, elderly people, pregnant and lactating women (indeed, women in general), culturally and linguistically diverse people, and people with disabilities are still underrepresented in clinical research.[125-127] An additional factor is failure to identify underrepresented people (e.g., transgender people) among recruited participants.[128] Ideally studies should be powered to allow analysis of major sub-group effects.[124,126]

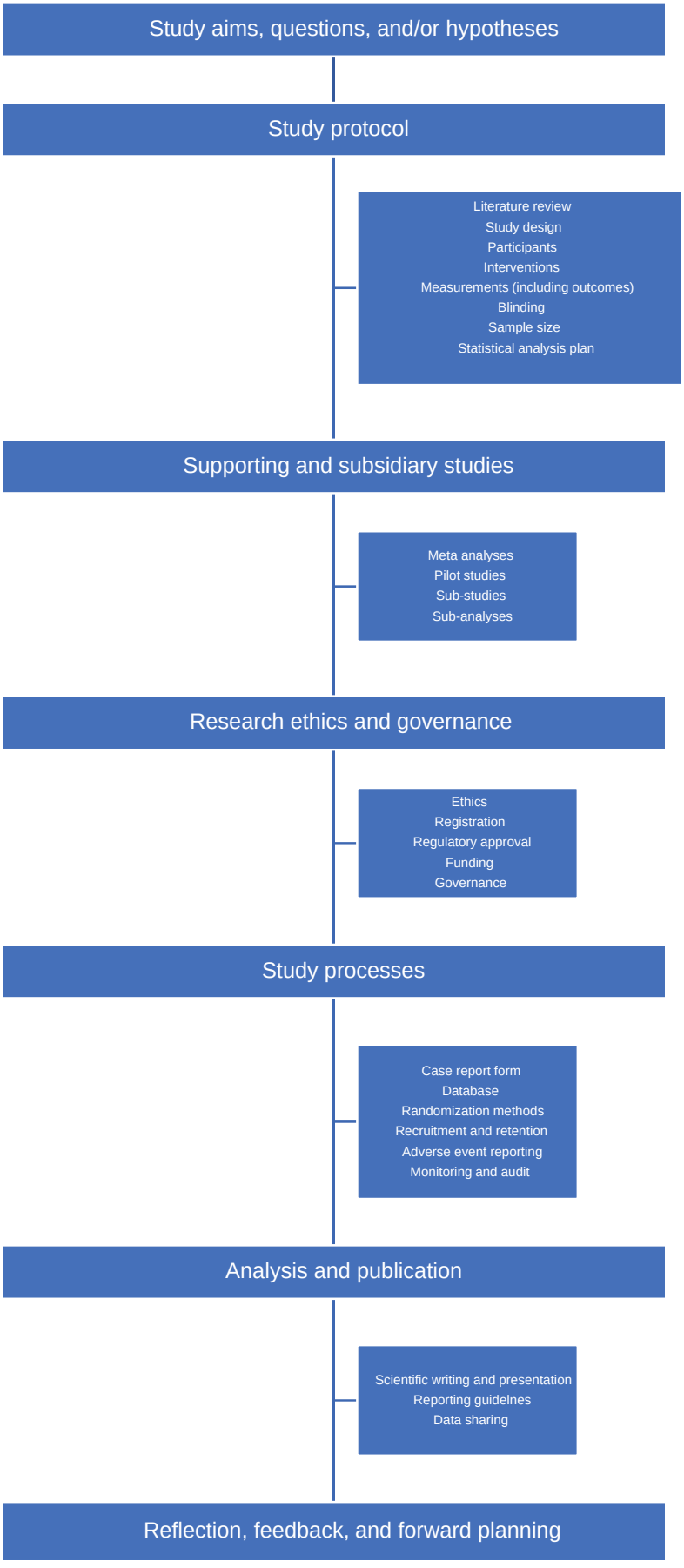

Fig. 89.5 Life cycle of a clinical research project.

INTERVENTIONS AND COMPARATORS

Nearly all clinical research involves an intervention (e.g., a drug, device, procedure, diagnostic test, bundle of care, etc.). In observational studies, the intervention is administered as part of usual care, whereas in experimental studies the intervention is administered by or on behalf of the investigators. Many studies involving an intervention also involve a comparator (e.g., placebo, usual care, or current best practice).[129] The comparator should be the best proven intervention, except when no such intervention exists.[109] In this case, a placebo may be appropriate. Usual care needs to be defined at the outset and monitored for change over time. Current best practice needs to be supported by guidelines and ensured during the study.[129] For a study to be ethical, there must be uncertainty in the expert community about the relative merits of the intervention(s) and comparator(s), and the study must be designed to resolve it (clinical equipoise).[130] Complex perioperative interventions and comparators may not be fully or properly implemented in pragmatic studies, and this variation may be nonrandom. For this reason, patients are usually analyzed in the groups to which they are assigned (intention-to-treat analysis).

OUTCOMES

Outcomes are events that may result from or be associated with an action taken during a study. Ideally, outcomes in clinical research influence the way clinicians practice and the choices that patients make.[131] Although patient-reported outcomes are subjective, they can greatly enhance the impact of the study if properly developed and implemented.[132,133] Binary outcomes that can be expressed as probabilities are easier to interpret than continuous outcomes, which can be converted to binary outcomes by defining an appropriately validated cut-off value.[131] Composite outcomes can be a useful means of assessing the impact of an event or intervention but can be misleading if the components are dissimilar in frequency, magnitude, direction of effect, or importance.[134] Surrogate outcomes occur between the event or intervention and the true outcome, and are used to draw conclusions about the true outcome. To be valid, the relationship between the surrogate and true outcome must be precisely defined.[135] Occasionally surrogate outcomes indicate the direction and magnitude of the effect of the intervention on the true outcome (e.g., the effect of anti-arrhythmic drugs on ventricular ectopic beats and sudden death post-myocardial infarction[136]). The primary outcome is the focus of the study and will be the basis of the sample size calculation. Secondary and safety outcomes should reflect the important beneficial and adverse effects of the event or intervention. Careful definition of outcomes and the timing of their measurement are fundamental to protocol development.

SAMPLE SIZE

A study needs a sufficient number of participants to provide reliable conclusions about the specified outcomes and treatment effects. If the number of study participants is too low, the investigators could incorrectly conclude that there is no treatment effect.[137] If the number of study participants is too large, it may delay the availability of information that is vital to patient care. In both cases resources will be wasted and participants will be exposed to unnecessary risk. A sample size calculation therefore should be conducted in the design phase of all studies. In qualitative and observational research, investigators should justify the reasons for their sampling frame (the population from which the sample is drawn).[20] This should be based on the availability of interviewees or relevant data in an appropriate format, or the investigators' expectations about the numbers of eligible participants and/or their estimates of acceptable 95% confidence intervals around the incidence of the primary outcome.[138] In comparative studies, sample size calculations are based on the expected difference in the primary outcome between the groups (effect size), the variance in the effect size (for continuous variables), and the risk that the investigators are willing to accept of false positive (α; type 1 error) and false negative (β; type 2 error) findings.[139,140] Effect sizes and variances can be estimated from the literature, pilot studies, statistical methods, or opinions about minimal clinically important differences.[141] The planned sample size will also depend on the number of groups, number of anticipated dropouts, and statistical analyses planned. The protocol should provide sufficient information to allow replication of quantitative sample size calculations.[138,140] In the anesthesia literature this is frequently not the case.[142] Post hoc calculation of statistical power using the results of a trial is inappropriate; the width of the confidence intervals around the primary outcome is a better indicator of the reliability of the result.[140]

DATA ANALYSIS PLAN

The details of data analysis are outside the scope of this chapter, and increasingly outside the scope of clinician researchers and reviewers.[143] Statistics training or collaboration with biostatisticians or experts in qualitative data analysis is commonly required to reach the level of statistical excellence required by regulators, funders, and journals. Statistical input is vital during the planning of clinical research, particularly in relation to the sample size calculation. The protocol should also include a plan for description of the data and analyses of the primary and secondary outcomes, subgroup and adjusted analyses, sensitivity analyses, interim analyses and stopping rules, and protocol adherence, as applicable.[115,116,144-146] The detailed statistical analysis plan for large observational studies and clinical trials is often published before the unblinded data are available, and at least should be predefined.[147] Peer-reviewed journals may require submission of the statistical analysis plan with the manuscript, require statistical checklists, and employ statistical editorial teams to assist with the evaluation of manuscripts.[143,146]

SUPPORTING STUDIES

Feasibility and Pilot Studies

Most clinical studies are inspired by prior research of some kind. Increasingly, however, preliminary work is specifically undertaken to demonstrate the viability of a future large observational or experimental study.[148,149] A feasibility study tests whether a future study can be done,[150]

examining the knowledge and interest of practitioners, availability of eligible patients, their willingness to participate, ease of applying the intervention and collecting data, qualities of the primary outcome, and resource requirements of the study. A pilot study is a type of feasibility study that tests the proposed hypotheses of a future study, without being scaled to test the effectiveness of interventions nor the strength of associations.[130] Pilot studies are frequently used to inform sample size calculations for the future studies, although this process may be flawed.[151] The rationale for the sample size of the feasibility study itself should be congruent with the feasibility objectives, but need not involve a formal sample size calculation.[150] Pilot studies are subject to the same ethical and regulatory requirements as definitive studies.

Sub-Studies

Sub-studies are an efficient means of investigating additional research questions using additional data collected in subsets of participants in large clinical trials. They are subject to the same ethical and regulatory requirements as the main study, and ideally are planned simultaneously. Sub-studies can investigate additional outcomes related to the randomized intervention, in which case they can be considered nested randomized trials.[152,153] They can investigate the associations of additional unrandomized exposures (e.g., biomarkers) with the same outcomes, in which case they can be considered nested cohort studies.[152,154] Finally, they can investigate the associations of additional unrandomized exposures with additional outcomes: these are also nested cohort studies and are an efficient use of a unique cohort.[152,155] The effect of additional randomized exposures is more properly evaluated using a factorial study design. Principal design considerations for sub-studies include adequate sample size and limiting the burden on investigators and patients.

Sub-Analyses

Sub-analyses are an efficient means of investigating additional research questions using data collected for another primary research purpose. Some sub-analyses are planned before the main study begins (e.g., population subgroup analyses). Others are planned after the data have been collected or the main results have been published and may address unanticipated events or discoveries. In any case the statistical analysis plan should be finalized before the analyses begin.[147] Sub-analyses can investigate the associations of unrandomized exposures (e.g., baseline characteristics, processes of care, measured variables) with the primary and secondary outcomes,[156,157] or the detailed effects of the randomized intervention on secondary outcomes.[158,159] Propensity scoring methods are increasingly used to determine the probability of a participant receiving a non-randomized treatment given a particular set of baseline characteristics, and to compensate for this by stratifying, matching, weighting, or adjusting on the basis of the propensity score.[156,160,161] However these methods are heavily reliant on the collection of appropriate and complete baseline data, and do not reduce confounding by unmeasured or unknown factors (e.g., the reasons the anesthesiologist chose a particular technique or defended a particular blood pressure).[160,161]

Ethical and Regulatory Considerations

ETHICS APPROVAL

All research on humans must be conducted within a system that ensures the safety and privacy of the participants.[109,113,162-166] This system can be local, national, or multinational. The degree of scrutiny is based on the potential for risk, discomfort, inconvenience, burden, and threats to privacy. Depending upon the jurisdiction, low-risk research (e.g., audits and surveys) may be conducted without review under policies that promote ethical conduct, or under expedited review processes.[164,166] Research that is not low-risk is considered by an institutional review board. The board approves the processes for obtaining consent or may waive or qualify consent if the risk to participants is low.[109,113,162-165] Deferred consent is also possible in urgent care research in some jurisdictions.[166] Approaching patients for the first time about a study on the day of surgery is acceptable to patients,[167,168] but participation rates are lower than before the day of surgery.[169]

REGISTRATION

Clinical trial registration was introduced in response to concerns about publication bias and selective reporting.[170] Additional aims of registration include reducing waste from unnecessary duplication of studies and improving access to clinical trial participation and results for patients.[171] Initial efforts involved mandating registration of a minimum dataset of protocol information before the enrollment of the first patient.[117,172] Subsequently efforts were expanded to mandating public disclosure of aggregated trial results, including those that were negative or inconclusive.[16,109,172-174] These efforts have not been completely successful.[171] For example, in a report about anesthesiology registration, although rates of registration had improved since 2007, 62% of clinical trials published in six specialty journals in 2015 were still inadequately registered.[175] Further culture change among funders, institutional review boards, investigators, and publishers will be required.[171,176] Registration of observational research is encouraged,[177] but is currently voluntary, partly because of concerns that registration may stifle exploratory analyses.[178]

REGULATORY APPROVAL

Federal, national, and state government agencies regulate the manufacture, import/export, supply, marketing, and surveillance of drugs and medical devices, with the aim of optimizing access to safe and effective therapeutic goods. There is some harmonization between jurisdictions.[179,180] Regulatory authorization is required for clinical trials of therapeutic goods that are not yet approved (phase I-III trials) or are being tested for an indication outside the current approval (phase IV trials).[181] The sponsor, institutional review board, and regulator collaborate in protecting patients who are exposed to unapproved therapeutic goods during clinical trials. This collaboration can be especially complex for international investigator-initiated clinical trials, and must be carefully considered in the planning phase.

DATA SHARING

Sharing of individual patient data from clinical trials is in the public interest, on scientific, economic, and ethical grounds.[182] Third parties may wish to verify trial results, correct errors, explore new hypotheses, or use the data in individual patient data metaanalyses.[175] Funding agencies, publishers, investigators, and the pharmaceutical industry have issued position statements about data sharing.[182-185] Retreating from its initial proposal of mandating data sharing,[186] the International Committee of Medical Journal Editors currently requires investigators to include a data sharing plan in the trial's registration and a data sharing statement in the primary manuscript.[184] These statements and plans must disclose whether, what, with whom, and how data will be shared. This position recognizes the current lack of policies, resources, and culture that will protect the interests of patients and investigators.[184] In order to be effective and responsible, data sharing must be planned from the outset of a clinical trial, as it involves clarifying the need for patient consent, constructing appropriate data management systems, and securing adequate funding.[185,187]

Study Management

FINANCIAL MANAGEMENT

Good clinical practice requires investigators and sponsors to document and agree on the financial aspects of a study.[113] Budgets and contracts are approved as part of the research governance process. All clinical research has a cost: even a case report requires retrieval of medical records, preparation of illustrations, and investment of investigator time. The cost of research is increasing in line with the increased size and complexity of clinical research. At the same time, the ability or inclination of health services to absorb costs that are not directly related to patient care is decreasing. Investigators therefore must obtain funding from other sources, including government agencies, commercial enterprises, and philanthropic organizations. This process can be time-consuming and stressful.[188] One study estimated that 34 working days of the principal writer's time was spent preparing each new proposal for a national funding round.[189] Only 20% of these proposals were funded. Streamlined and flexible application processes, and rules regarding resubmission of unsuccessful proposals, may reduce workload and increase success.[188,189]

DATA MANAGEMENT

The protocol should list each item of data to be collected, its source, and the timing of measurement.[115] The ethics review process addresses whether the proposed data collection meets privacy and data security requirements.[113] In explanatory studies, such as phase I-III clinical trials, the volume of data collected can be large, whereas in pragmatic studies data collection should be limited to vital measurements only.[190] Three main options are available for data collection: (1) a bespoke case report form; (2) data extraction from existing sources (such as medical records, registries, and administrative databases); and (3) a hybrid approach.[190] Case report forms can be printed or electronic, with electronic forms offering mechanisms to ensure complete and accurate data entry.[191] Data will then be transferred to a database that is usually established specifically for the study. Increasingly anesthesia research protocols require linkage of records of individual patients or groups across unrelated databases, requiring the input of health informaticians and raising issues of privacy and data security.[192]

HUMAN RESOURCE MANAGEMENT

The human resources associated with clinical research include the investigators, trial coordinators, treating clinicians, and patients. Ethics and governance processes address whether the legitimate interests and rights of these parties are protected during the study.[113] Investigators and trial coordinators should be suitability qualified, experienced, and competent for the research they propose to undertake, and ideally should have International Council for Harmonisation E6 Good Clinical Practice certification[113] (a requirement of some funding bodies[193]). Clinical trial networks can play a big role in nurturing careers and fostering collaboration. Treating clinicians need to be advised about their role in implementing the protocol and be engaged in ensuring the success of the study. This is especially important in protocols that require the continuous attention of the treating clinician during an anesthetic[194] or critical care unit stay.[85] Recruiting patients to anesthesia and critical care studies can be difficult, because of time constraints and multiple competing priorities. A systematic review of recruitment strategies identified that open rather than blinded treatment allocation, and phone follow-up to written invitations to participate, significantly improved recruitment.[195] Loss to follow-up is very low in anesthesia and critical care studies, because time frames for data collection are often short and patients are in the operating room or hospital for most or all of the study.[85,156,194] A systematic review of retention strategies in ambulatory settings reported that only monetary incentives improved retention.[196]

ADVERSE EVENT REPORTING

Adverse event reporting is a crucial step in ensuring the safety of participants in all clinical research (Box 89.3).[197] An adverse event is any unfavorable and unintended occurrence associated with the use of a medicinal product, whether or not it is related to that product. Adverse events can be categorized by severity (intensity), seriousness (effect on patient outcome), expectedness (observed before), and causality (attributability to the medicinal product).[197] Expedited reporting of adverse events to regulators and institutional review boards is an essential part of phase I to III trials.[172,198] For phase IV trials, universal safety endpoints and unexpected adverse events are reported periodically by the trial's data and safety monitoring board. Only serious unexpected events that would be reported as part of routine clinical care are reported immediately to regulators and institutional review boards. The primary report of a clinical trial should include a table of adverse events by system and/or severity.[172,198] Adverse event reporting is burdensome and costly, and therefore should be aligned with the risk to patient safety of the medicinal product.[199-201]

BOX 89.3 Definitions of Adverse Events in Clinical Research

Adverse event: Any untoward medical occurrence associated with the use of a drug in humans, whether or not considered drug related.

Suspected adverse reaction: Any adverse event for which there is a reasonable possibility that the drug caused the adverse event.

Unexpected: Any adverse event or suspected adverse reaction that is not listed in the investigator brochure (or general investigational plan) or is not listed at the specificity or severity that has been observed.

Serious adverse event or suspected adverse reaction: Any adverse event or suspected adverse reaction that results in any of the following outcomes: death, a life-threatening adverse event, inpatient hospitalization or prolongation of existing hospitalization, a persistent or significant incapacity or substantial disruption of the ability to conduct normal life functions, or a congenital anomaly/birth defect.

Life-threatening: An adverse event or suspected adverse reaction that places the patient or subject at immediate risk of death. It does not include an adverse event or suspected adverse reaction that, had it occurred in a more severe form, might have caused death.

MONITORING AND AUDIT

Monitoring is "the act of overseeing the progress of a clinical trial."[113] All clinical studies require some monitoring, to ensure that participants are protected; the study is being conducted in accordance with its approvals; and the data are complete, accurate, and verifiable.[113] Monitoring may range from simple checks by the investigators through on-site monitoring by the sponsor or a contract research organization. Increasingly, centralized monitoring is being used to look for data trends and integrity issues.[202] Audit is "a systematic and independent examination of trial related activities and documents."[113] Audits are conducted by institutional review boards, regulators, and funders, sometimes after a concern is raised about a study. In clinical trials, quality management is the responsibility of the data and safety monitoring board. The board can review unblinded data and adverse event reports, and make assessments about the evolving balance between efficacy and risk in the study. Boards and board members need to be independent, and increasingly need to be trained for the heavy academic, legal, and ethical responsibilities involved.[203]

Study Reporting

MANUSCRIPT

Accurate, complete, and timely reporting of results supports research integrity and usefulness. Reporting guidelines are available for a wide range of study designs and populations, and compliance with these guidelines and completion of checklists and flow diagrams is mandated by some journals.[117,204] For example, the STROBE and Consolidated Standards Of Reporting Trials (CONSORT) statements aim to improve the quality of reporting of observational studies and randomized trials respectively.[20,138-140] The quality of reporting of randomized controlled trials has improved over time, but it is unclear if this is due to the CONSORT statement.[205] A systematic review about the effect of journal endorsement of the statement on quality of reporting identified a significant improvement in reporting for only five of the 27 items,[206] highlighting the need for research into adherence to guidelines.[207] Research is only useful if the investigators communicate effectively with their intended audience. As well as reporting guidelines[204] and instructions from journal editors,[117] researchers are advised to consult resources about writing clear and graceful prose.

AUTHORSHIP

Investigators can be involved in the concept, design, funding, governance, management, data collection, analysis, and/or reporting of clinical research. The International Committee of Medical Journal Editors recommends that authorship be based on substantial contributions to the research, approval of the submitted manuscripts, and agreement to be accountable for all aspects of the work.[117] Honorary authorship to recognize contributions that do not meet the recommendations (such as leading the department in which the work was done or technical editing of the manuscript) is thereby prohibited. Criteria for authorship and for acknowledging non-author contributions should be established before the research commences and reviewed at manuscript submission.[117,208] With increases in numbers of authors, multi-investigator groups, and data sharing, solutions to acknowledging contributions to research are rapidly evolving.[208-210]

PUBLISHING

Researchers need to choose their target journal with care, taking into consideration the subject and potential impact of their research, the requirements of funders for open access publication, and the bona fides of the target journal. Legitimate publishers conform to recognized standards of publication ethics.[117,211] Open access options are available for a fee in traditional subscription journals and legitimate open access journals. So-called "predatory" open access journals aggressively seek submissions through "spam" email, do not conform to ethical standards, and may charge high fees without delivering on open access or indeed publishing at all.[212,213] Various resources are available to help authors identify predatory journals, although these are imperfect due to increasing numbers of both legitimate and predatory journals.[212-214]

RESEARCH INTEGRITY

Research integrity is "active adherence to the ethical principles and professional standards essential for the responsible practice of research."[215] Research misconduct includes fabrication, falsification, plagiarism, misleading reporting, redundant publication, abuse of authorship, conflict of interest infringements, funding infringements, fraud in the peer review process, and unethical research (in particular, in relation to patient consent and protection).[216-218] Research misconduct damages the reputation of science and scientists,

and is a threat to patient safety. A systematic review on interventions to improve research integrity concluded that the evidence for effectiveness was sparse, of low quality, and concentrated on short-term outcomes.[219] Plagiarism, which ranges from intellectual theft through careless practice to self-plagiarism (not strictly speaking plagiarism but possibly duplicate publication),[220] is the most common form of research misconduct.[221] Practical training may reduce plagiarism[219] and routine screening using text-matching software may reduce plagiarism in submitted papers.

Reflection, Feedback, and Forward Planning

Continuous quality improvement is now mainstream in clinical practice and medical education.[222] Practitioners and students are encouraged to reflect on their experiences; note what they did well and where they could do better; discuss these reflections with their supervisors or teachers; and give their colleagues or teachers feedback about the quality of the workplace or learning environment. At the end of this cycle plans are made for the future. Researchers could benefit from a similar cycle of reflection, feedback, and forward planning. While a lot of helpful information and training about designing and conducting clinical research is available, including in *Miller's Anesthesia*, each researcher has unique aptitudes and works in a unique context. We recommend that researchers keep a journal about each research project, reflect on their experience when the project is completed, and share strategies for success with their peers and junior colleagues. Finally, as well as giving patients an opportunity to contribute to study design, giving patients an opportunity to comment on the results of studies in which they participated might give researchers fresh insights and open up new avenues of research.

Complete references available online at expertconsult.com.

References

1. Institute of Medicine. Washington: National Academy Press; 2001.
2. Coggon D, et al. http://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated.
3. Keus F, et al. *BMC Med Res Methodol*. 2010;10:90.
4. Schwartz S, et al. *Epidemiology*. 2015;26:216.
5. Sessler DI, Imrey PB. *Anesth Analg*. 2015;121:1034.
6. Vetter TR, Mascha EJ. *Anesth Analg*. 2017;125.
7. Short T, Leslie K. *Br J Anaesth*. 2014;113:897.
8. Chu R, et al. *PLoS One*. 2012;7:e36677.
9. Zhao W. *Contemp Clin Trials*. 2013;36:263.
10. Moons KG, et al. *BMJ*. 2009;338:b606.
11. Song F, et al. *Health Technol Assess*. 2010;14:1.
12. Akl EA, et al. *BMJ Open*. 2015;5:e008431.
13. Wasserstein R, Lazar N. *Am Stat*. 2016;70:129.
14. Wijeysundera DN, et al. *J Clin Epidemiol*. 2009;62:13. e15.
15. Ioannidis JPA. *JAMA*. 2018 (in press).
16. World Health Organization. International clinical trials registry platform. international standards for clinical trial registries. http://apps.who.int/iris/bitstream/10665/76705/1/9789241504294_eng.pdf?ua=1&ua=1.
17. Sessler DI, Imrey PB. *Anesth Analg*. 2015;121:1043.
18. Doll R, Hill AB. *BMJ*. 1950;2:739.
19. Hill AB. *Proc R Soc Med*. 1965;58:295.
20. von Elm E, et al. *PLoS Med*. 2007;4:e296.
21. Rao A, et al. *PLoS One*. 2016;11:e0155078.
22. Kruger BD, et al. *Anesth Analg*. 2017;125:1898.
23. Rosenblatt MA, et al. *Anesthesiology*. 2006;105:217.
24. Denborough MA, et al. *Br J Anaesth*. 1962;34:395.
25. Griffith H, Johnson E. *Anesthesiology*. 1942;3:418.
26. Brain AI. *Br J Anaesth*. 1983;55:801.
27. Rodseth RN, et al. *J Am Coll Cardiol*. 2011;58:522.
28. Moons KG, et al. *Clin Chem*. 2004;50:473.
29. O'Leary JD, et al. *Anesthesiology*. 2016;125:272.
30. McIsaac DI, et al. *Anesthesiology*. 2018;128:480.
31. Riazi S, et al. *Diabetes Care*. 2012;35:2575.
32. Chung F, et al. *Anesthesiology*. 2008;108:812.
33. Busato P, et al. *Sao Paulo Med J*. 2017;135:116.
34. Bijker JB, et al. *Anesthesiology*. 2012;116:658.
35. Nuttall GA, et al. *Anesth Analg*. 2001;93:1410.
36. Wacholder S, et al. *Am J Epidemiol*. 1992;135:1019.
37. Devereaux PJ, et al. *JAMA*. 2017;317:1642.
38. Wijeysundera DN, et al. *Lancet*. 2008;372:562.
39. Liu JB, et al. *Anesthesiology*. 2018;128:283.
40. Ladha K, et al. *BMJ*. 2015;351:h3646.
41. Rubin DB. *Stat Med*. 2007;26:20.
42. Minto C, et al. *Anesthesiology*. 1997;86:10.
43. Schnider T, et al. *Anesthesiology*. 1999;90:1502.
44. Kannampallil T, et al. *Pain*. 2016;157:2739.
45. Ma Y, et al. *Reg Anesth Pain Med*. 2012;37:99.
46. Reynolds PS, et al. *Anesth Analg*. 2016;123:1400.
47. Shulman M, et al. *Anesthesiology*. 2015;122:524.
48. Abu-Arafeh A, et al. *Br J Anaesth*. 2016;117:569.
49. Fitzpatrick R, et al. *Health Technol Assess*. 1998;2(i–iv):1.
50. Hunt TD, et al. *BMJ Open*. 2012;2:e000665.
51. Hooper TD, et al. *Anaesth Intensive Care*. 2015;43:461.
52. Wijeysundera DN. *Can J Anaesth*. 2016;63:148.
53. Rathore SS, et al. *Circulation*. 2003;107:811.
54. Kalkman CJ, et al. *Pain*. 2003;105:415.
55. Sheth T, et al. *BMJ*. 2015;350:h1907.
56. Goldman L, et al. *N Engl J Med*. 1982;307:588.
57. Canet J, et al. *Anesthesiology*. 2010;113:1338.
58. Bilimoria KY, et al. *J Am Coll Surg*. 2013;217:833.
59. Zimmerman JE, et al. *Crit Care Med*. 2006;34:1297.
60. American Statistical Association. What is a survey? www.amstat.org/sections/srms/whatsurvey.html.
61. Jones D, et al. *Anaesth Intensive Care*. 2006;34:245.
62. Tait AR, Voepel-Lewis T. *Paediatr Anaesth*. 2015;25:656.
63. Klabunde CN, et al. *Eval Health Prof*. 2012;35:477.
64. Story DA, et al. *Anesth Analg*. 2011;113:591.
65. Lohr KN, Steinwachs DM. *Health Services Research*. 2002;37:7.
66. Wijeysundera DN, et al. *Anesthesiology*. 2012;116:25.
67. Moonesinghe SR, et al. *BMJ Open*. 2017;7:e017690.
68. Lane-Fall MB, et al. *BMC Surgery*. 2014;14:96.
69. Emond YE, et al. *Implement Sci*. 2015;10:3.
70. Harris AD, et al. *Clin Infect Dis*. 2004;38:1586.
71. Sterne JA, et al. *BMJ*. 2016;355:i4919.
72. Paneth N, Fine P. *Lancet*. 2013;381:1267.
73. Ho AMH, et al. *Anesth Analg*. 2017.
74. Alvarez MP, et al. *Surg Endosc*. 2015;29:2506.
75. Streeter AJ, et al. *J Clin Epidemiol*. 2017;87:23.
76. Wellek S, Blettner M. *Dtsch Arztebl Int*. 2012;109:276.
77. Niesters M, et al. *Br J Anaesth*. 2013;110:1010.
78. Colloca L, Miller FG. *Psychosomatic Med*. 2011;73:598.
79. Sackett DL, Rosenberg WM. . *J R Soc Med*. 1995;88:620.
80. Roberts C, Torgerson D. *BMJ*. 1998;317:1301.
81. Suresh K. *J Hum Reprod Sci*. 2011;4:8.
82. Torgerson DJ, Roberts C. *BMJ*. 1999;319:375.
83. Wartolowska K, et al. *F1000Res*. 2017;6:1663.
84. Candiotti KA, et al. *Anesth Analg*. 2008;107:445.
85. Finfer S, et al. *N Engl J Med*. 2004;350:2247.
86. van Wijck AJ, et al. *Lancet*. 2006;367:219.
87. Mascha EJ. *Anesthesiology*. 2010;113:779.
88. Davidson AJ, et al. *Lancet*. 2016;387:239.
89. Connolly SJ, et al. *N Engl J Med*. 2009;361:1139.
90. Sessler DI, Imrey PB. *Anesth Analg*. 2015;121:1052.
91. Apfel C, et al. *N Engl J Med*. 2004;350:2441.
92. Myles PS, et al. *N Engl J Med*. 2017;376:136.
93. Myles PS, et al. *N Engl J Med*. 2016;374:728.

94. Brown C, Lilford R. *BMJ*. 2008;337:a2764.
95. de Smet AM, et al. *N Engl J Med*. 2009;360:20.
96. Hillman K, et al. *Lancet*. 2005;365:2091.
97. Hemming K, et al. *BMJ*. 2015;350:h391.
98. Ellenberg SS. *JAMA*. 2018;319:607.
99. The Gambia Hepatitis Study Group. *Cancer Res*. 1987;47:5782.
100. Moller AM, Myles PS. *Br J Anaesth*. 2016;117:428.
101. Moher D, et al. *PLoS Med*. 2009;6:e1000097.
102. Pussegoda K, et al. *Syst Rev*. 2017;6:131.
103. Sivakumar H, Peyton PJ. *Br J Anaesth*. 2016;117:431.
104. The Cochrane Collaboration. RevMan5. London. http://www.cochrane.org.
105. The Cochrane Collaboration. Covidence. London. http://community.cochrane.org/tools/review-production-tools/covidence.
106. Debray TP, et al. *Res Synth Methods*. 2015;6:293.
107. Thomas D, et al. *BMC Med Res Methodol*. 2014;14:79.
108. Nevitt SJ, et al. *BMJ*. 2017;357:j1390.
109. World Medical Association. www.wma.net/en/30publications/10policies/b3/index.html. Accessed March 8 2018.
110. Pildal J, et al. *BMJ*. 2005;330:1049.
111. Hrobjartsson A, et al. *J Clin Epidemiol*. 2009;62:967.
112. World Health Organization. http://www.who.int/rpc/research_ethics/format_rp/en/. Accessed April 8 2018.
113. ICH Expert Working Group. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R2__Addendum_Step2.pdf. Accessed April 8 2018.
114. National Institutes of Health and Food and Drug Agency. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-064.html. Accessed March 6 2018.
115. Chan AW, et al. *Ann Intern Med*. 2013;158:200.
116. Chan AW, et al. *BMJ*. 2013;346:e7586.
117. International Committee of Medical Journal. In: *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals*. 2017. http://www.icmje.org/icmje-recommendations.pdf. Accessed April 7 2018.
118. Rios LP, et al. *BMC Med Res Methodol*. 2010;10:11.
119. Haynes RB. *J Clin Epidemiol*. 2006;59:881.
120. Nizamuddin SL, et al. *J Cardiothorac Vasc Anesth*. 2017;31:1878.
121. Loudon K, et al. *BMJ*. 2015;350:h2147.
122. Oude Rengerink K, et al. *J Clin Epidemiol*. 2017;89:173.
123. Sackett DL. *Pol Arch Med Wewn*. 2011;121:259.
124. National Institutes of Health. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-18-014.html.
125. Geller SE, et al. *J Womens Health (Larchmt)*. 2011;20:315.
126. Spong CY, Bianchi DW. *JAMA*. 2018;319:337.
127. Hughson JA, et al. *Trials*. 2016;17:263.
128. Reisner SL, et al. *Lancet*. 2016;388:412.
129. Zuidgeest MGP, et al. *J Clin Epidemiol*. 2017;90:92.
130. Freedman B. *N Engl J Med*. 1987;317:141.
131. Welsing PM, et al. *J Clin Epidemiol*. 2017;90:99.
132. Kyte D, et al. *PLoS One*. 2014;9:e110229.
133. Calvert M, et al. *JAMA*. 2018;319:483.
134. Myles PS, Devereaux PJ. *Anesthesiology*. 2010;113:776.
135. Baker SG. *Stat Med*. 2018;37:507.
136. The CAST Investigators. *N Engl J Med*. 1989;321:406.
137. Halpern SD, et al. *JAMA*. 2002;288:358.
138. Vandenbroucke JP, et al. *PLoS Med*. 2007;4:e297.
139. Schulz KF, et al. *Trials*. 2010;11:32.
140. Moher D, et al. *J Clin Epidemiol*. 2010;63:e1–37.
141. Cook JA, et al. *Health Technol Assess*. 2014;18:1.
142. Abdulatif M, et al. *Br J Anaesth*. 2015;115:699.
143. Mascha EJ, Vetter TR. *Anesth Analg*. 2017;124:719.
144. ICH Expert Working Group. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf.
145. Thomas L, Peterson ED. *JAMA*. 2012;308:773.
146. Lang TA, Altman DG. *Int J Nurs Stud*. 2015;52:5.
147. Eisenach JC, et al. *Anesthesiology*. 2016;124:998.
148. Thabane L, et al. *Trials*. 2011;12:48.
149. Story D. et al. *Anaesth Intensive Care*. 2018.
150. Eldridge SM, et al. *BMJ*. 2016;355:i5239.
151. Kraemer HC. et al. *Arch Gen Psychiatry*. 2006;63:484.
152. Kirkegaard H, et al. *JAMA*. 2017;318:341.
153. Bro-Jeppesen J, et al. *Circ Cardiovasc Interv*. 2014;7:663.
154. Frydland M, et al. *Am J Cardiol*. 2016;118:998.
155. Gilje P, et al. *Resuscitation*. 2016;107:156.
156. Devereaux P, et al. *N Engl J Med*. 2014;370:1504.
157. Sessler DI, et al. *Anesthesiology*. 2018;128:317.
158. Devereaux P, et al. *N Engl J Med*. 2014;370:1494.
159. Eikelboom JW, et al. *Anesthesiology*. 2016;125:1121.
160. Elze MC, et al. *J Am Coll Cardiol*. 2017;69:345.
161. Leslie K, et al. *Br J Anaesth*. 2013;111:382.
162. Department of Health and Human Services. https://www.hhs.gov/ohrp/regulations-and-policy/regulations/index.html. Accessed April 7 2018.
163. Canadian Institutes of Health Research. http://www.pre.ethics.gc.ca/pdf/eng/tcps2-2014/TCPS_2_FINAL_Web.pdf. Accessed April 7 2018.
164. Health Research Authority. https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/uk-policy-framework-health-social-care-research/. Accessed March 8 2018.
165. Australian Government. https://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research. Accessed April 7 2018.
166. European Commission. https://ec.europa.eu/health/sites/health/files/files/clinicaltrials/2012_07/proposal/2012_07_proposal_en.pdf. Accessed April 7 2018.
167. Australian Government. https://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/e111_ethical_considerations_in_quality_assurance_140326.pdf. Accessed April 7 2018.
168. Murphy GS, et al. *Anesthesiology*. 2016;124:1246.
169. Chludzinski A, et al. *Mayo Clin Proc*. 2013;88:446.
170. Simes RJ. *J Clin Oncol*. 1986;4:1529.
171. Zarin DA, et al. *N Engl J Med*. 2017;376:383.
172. US Department of Health and Human Services. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM227351.pdf. Accessed April 7 2018.
173. European Federation of Pharmaceutical Industries and Associations. https://www.ifpma.org/wp-content/uploads/2010/11/Joint-Position-on-Disclosure-of-CT-Info-via-CT-Registries-Revised-Oct2017-vF.pdf. Accessed April 8 2018.
174. European Commission. Commission Guideline. https://ec.europa.eu/health//sites/health/files/files/eudralex/vol-10/2012_302-03/2012_302-03_en.pdf. Accessed April 8 2018.
175. Jones PM, et al. *Anesth Analg*. 2017;125:1292.
176. Nizamuddin J, et al. *Anesth Analg*. 2017;125:1098.
177. Loder E, et al. *BMJ*. 2010;340:c950.
178. Sorensen HT, Rothman KJ. *BMJ*. 2010;340:c703.
179. European Medicines Agency. Clinical trials in human medicines. http://www.ema.europa.eu/ema/index.jsp?curl=pages/special_topics/general/general_content_000489.jsp&mid=WC0b01ac058060676f. Accessed March 6 2018.
180. ICH Expert Working Group. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q11/Q11_Step_4.pdf. Accessed April 8 2018.
181. Absalom AR, et al. *Anesth Analg*. 2016;122:70.
182. Institute of Medicine (IOM). http://nationalacademies.org/hmd/reports/2015/sharing-clinical-trial-data.aspx. Accessed March 6 2018.
183. Pharmaceutical Research and Manufacturers of America, European Federation of Pharmaceutical Industries and Associations. http://phrma-docs.phrma.org/sites/default/files/pdf/PhRMAPrinciplesForResponsibleClinicalTrialDataSharing.pdf. Accessed April 8 2018.
184. Taichman DB, et al. *BMJ*. 2017;357:j2372.
185. Ohmann C, et al. *BMJ Open*. 2017;7:e018647.
186. Taichman DB, et al. *N Engl J Med*. 2016;374:384.
187. Tudur Smith C, et al. *Trials*. 2017;18:319.
188. Herbert DL, et al. *BMJ Open*. 2014;4:e004462.
189. Herbert DL, et al. *BMJ Open*. 2013;3.
190. Meinecke AK, et al. Series: Pragmatic trials and real world evidence: paper 8. Data collection and management. *J Clin Epidemiol*. 2017;91:13.
191. Bellary S, et al. *Perspect Clin Res*. 2014;5:159.
192. Morgan DJ, Ho KM*Anaesth Intensive Care*. 2016;44:237.
193. National Institutes of Health. https://grants.nih.gov/policy/clinical-trials/good-clinical-training.htm. Accessed April 8 2018.
194. Pearse RM, Harrison DA, MacDonald N, et al. *JAMA*. 2014;311:2181–2190.

195. Treweek S, et al. *Cochrane Database Syst Rev*. 2018;2:MR000013.
196. Brueton VC, et al. *Cochrane Database Syst Rev*. 2013:MR000032.
197. ICH Expert Working Group. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2A/Step4/E2A_Guideline.pdf. Accessed April 8 2018.
198. ICH Expert Working Group. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2F/Step4/E2F_Step_4.pdf. Accessed April 8 2018.
199. Reith C, et al. *N Engl J Med*. 2013;369:1061.
200. Wallace S, et al. *Med J Aust*. 2016;204:231.
201. European Medicines Agency, Heads of Medicines Agencies. http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2017/08/WC500232767.pdf. Accessed April 8 2018.
202. Pogue JM, et al. *Clin Trials*. 2013;10:225.
203. Fleming TR, et al. *Clin Trials*. 2017;14:115.
204. EQUATOR Network. http://www.equator-network.org/. Accessed April 8 2018.
205. Dechartres A, et al. *BMJ*. 2017;357:j2490.
206. Turner L, et al. *Cochrane Database Syst Rev*. 2012;11:MR000030.
207. Blanco D, et al. *BMJ Open*. 2017;7:e017551.
208. Fontanarosa P, et al. *JAMA*. 2017;318:2433–2437.
209. Bierer BE, et al. *N Engl J Med*. 2017;377:402.
210. US National Library of Medicine. http://www.nlm.nih.gov/bsd/authors1.html. Accessed April 8 2018.
211. Committee on Publication Ethics (COPE). Core practices. Washington. https://publicationethics.org/core-practices. Accessed March 6 2018.
212. Shen C, Bjork BC. *BMC Med*. 2015;13:230.
213. Moher D, et al. *Nature*. 2017;549:23.
214. Directory of Open Access Journals. Accessed April 8 2018.
215. Korenman SG. https://ori.hhs.gov/education/products/ucla/chapter1/page02.htm. Accessed April 8 2018.
216. Office of Science and Technology Policy. 2000;65:76260–76264.
217. A consensus statement on research misconduct in the UK. *BMJ*. 2012;344:e1111.
218. Haug CJ. Peer-review fraud - hacking the scientific publication process. *N Engl J Med*. 2015;373:2393–2395.
219. Marusic A, et al. *Cochrane Database Syst Rev*. 2016;4:MR000038.
220. Shafer SL. *Anesth Analg*. 2016;122:1776–1780.
221. Moylan EC, et al. *BMJ Open*. 2016;6:e012047.
222. Horsley T, et al. *BMJ Open*. 2016;6:e010368.

References

1. Institute of Medicine (IOM). *Crossing the Quality Chasm: a new Health System for the 21st Century*. Washington: National Academy Press; 2001.
2. Coggon D, Rose G, Barker D. *Epidemiology for the Uninitiated (4th edition)*. http://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated.
3. Keus F, Wetterslev J, Gluud C, van Laarhoven CJ. Evidence at a glance: error matrix approach for overviewing available evidence. *BMC Med Res Methodol*. 2010;10:90.
4. Schwartz S, Campbell UB, Gatto NM, Gordon K. Toward a clarification of the taxonomy of "bias" in epidemiology textbooks. *Epidemiology*. 2015;26:216–222.
5. Sessler DI, Imrey PB. Clinical research methodology 1:study designs and methodologic sources of error. *Anesth Analg*. 2015;121:1034–1042.
6. Vetter TR, Mascha EJ. Bias, confounding, and interaction: lions and tigers, and bears, oh my! *Anesth Analg*. 2017;125:1042–1048.
7. Short T, Leslie K. Known unknowns and unknown unknowns: electroencephalographic burst suppression and mortality. *Br J Anaesth*. 2014;113:897–899.
8. Chu R, Walter SD, Guyatt G, et al. Assessment and implication of prognostic imbalance in randomized controlled trials with a binary outcome--a simulation study. *PLoS One*. 2012;7:e36677.
9. Zhao W. Selection bias, allocation concealment and randomization design in clinical trials. *Contemp Clin Trials*. 2013;36:263–265.
10. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606.
11. Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess*. 2010;14:1–193.
12. Akl EA, Shawwa K, Kahale LA, et al. Reporting missing participant data in randomised trials: systematic survey of the methodological literature and a proposed guide. *BMJ Open*. 2015;5:e008431.
13. Wasserstein R, Lazar N. The ASA's statement on p-values: context, process, and purpose. *Am Stat*. 2016;70:129–133.
14. Wijeysundera DN, Austin PC, Hux JE, Beattie WS, Laupacis A. Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials. *J Clin Epidemiol*. 2009;62:13–21.e15.
15. Ioannidis JPA. The proposal to lower P value thresholds to .005. *JAMA*. 2018
16. World Health Organization. International clinical trials registry platform. International standards for clinical trial registries. http://apps.who.int/iris/bitstream/10665/76705/1/9789241504294_eng.pdf?ua=1&ua=1. Accessed March 18 2018.
17. Sessler DI, Imrey PB. Clinical research methodology 2: observational clinical research. *Anesth Analg*. 2015;121:1043–1051.
18. Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. *BMJ*. 1950;2:739–748.
19. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58:295–300.
20. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med*. 2007;4:e296.
21. Rao A, Bruck K, Methven S, et al. Quality of reporting and study design of ckd cohort studies assessing mortality in the elderly before and after STROBE: a systematic review. *PLoS One*. 2016;11:e0155078.
22. Kruger BD, Kurmann J, Corti N, Spahn DR, Bettex D, Rudiger A. Dexmedetomidine-associated hyperthermia: a series of 9 cases and a review of thse literature. *Anesth Analg*. 2017;125:1898–1906.
23. Rosenblatt MA, Abel M, Fischer GW, Itzkovich CJ, Eisenkraft JB. Successful use of a 20% lipid emulsion to resuscitate a patient after a presumed bupivacaine-related cardiac arrest. *Anesthesiology*. 2006;105:217–218.
24. Denborough MA, Forster JF, Lovell RR, Maplestone PA, Villiers JD. Anaesthetic deaths in a family. *Br J Anaesth*. 1962;34:395–396.
25. Griffith H, Johnson E. The use of curare in general anesthesia. *Anesthesiology*. 1942;3:418–420.
26. Brain AI. The laryngeal mask--a new concept in airway management. *Br J Anaesth*. 1983;55:801–805.
27. Rodseth RN, Lurati Buse GA, Bolliger D, et al. The predictive ability of pre-operative B-type natriuretic peptide in vascular patients for major adverse cardiac events: an individual patient data meta-analysis. *J Am Coll Cardiol*. 2011;58:522–529.
28. Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. *Clin Chem*. 2004;50:473–476.
29. O'Leary JD, Janus M, Duku E, et al. A population-based study evaluating the association between surgery in early life and child development at primary school entry. *Anesthesiology*. 2016;125:272–279.
30. McIsaac DI, Wijeysundera DN, Huang A, Bryson GL, van Walraven C. Association of hospital-level neuraxial anesthesia use for hip fracture surgery with outcomes: a population-based cohort study. *Anesthesiology*. 2018;128:480–491.
31. Riazi S, Bril V, Perkins BA, et al. Can ultrasound of the tibial nerve detect diabetic peripheral neuropathy? A cross-sectional study. *Diabetes Care*. 2012;35:2575–2579.
32. Chung F, Yegneswaran B, Liao P, et al. STOP questionnaire: a tool to screen patients for obstructive sleep apnea. *Anesthesiology*. 2008;108:812–821.
33. Busato P, Garbin RR, Santos CN, Paranhos LR, Rigo L. Influence of maternal anxiety on child anxiety during dental care: cross-sectional study. *Sao Paulo Med J*. 2017;135:116–122.
34. Bijker JB, Persoon S, Peelen LM, et al. Intraoperative hypotension and perioperative ischemic stroke after general surgery: a nested case-control study. *Anesthesiology*. 2012;116:658–664.
35. Nuttall GA, Garrity JA, Dearani JA, Abel MD, Schroeder DR, Mullany CJ. Risk factors for ischemic optic neuropathy after cardiopulmonary bypass: a matched case/control study. *Anesth Analg*. 2001;93:1410–1416.
36. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I. Principles. *Am J Epidemiol*. 1992;135:1019–1028.
37. Devereaux PJ, Biccard BM, Sigamani A, et al. Association of postoperative high-sensitivity troponin levels with myocardial injury and 30-day mortality among patients undergoing noncardiac surgery. *JAMA*. 2017;317:1642–1651.
38. Wijeysundera DN, Beattie WS, Austin PC, Hux JE, Laupacis A. Epidural anaesthesia and survival after intermediate-to-high risk non-cardiac surgery: a population-based cohort study. *Lancet*. 2008;372:562–569.
39. Liu JB, Liu Y, Cohen ME, Ko CY, Sweitzer BJ. Defining the intrinsic cardiac risks of operations to improve preoperative cardiac risk assessments. *Anesthesiology*. 2018;128:283–292.
40. Ladha K, Vidal Melo MF, McLean DJ, et al. Intraoperative protective mechanical ventilation and risk of postoperative respiratory complications: hospital based registry study. *BMJ*. 2015;351:h3646.
41. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Stat Med*. 2007;26:20–36.
42. Minto C, Schnider T, Egan T, et al. Influence of age and gender on the pharmacokinetics and pharmacodynamics of remifentanil. I. Model development. *Anesthesiology*. 1997;86:10–23.
43. Schnider T, Minto C, Shafer S, et al. The influence of age on propofol pharmacodynamics. *Anesthesiology*. 1999;90:1502–1516.
44. Kannampallil T, Galanter WL, Falck S, et al. Characterizing the pain score trajectories of hospitalized adult medical and surgical patients: a retrospective cohort study. *Pain*. 2016;157:2739–2746.
45. Ma Y, Mazumdar M, Memtsoudis SG. Beyond repeated-measures analysis of variance: advanced statistical methods for the analysis of longitudinal data in anesthesia research. *Reg Anesth Pain Med*. 2012;37:99–105.
46. Reynolds PS, Middleton P, McCarthy H, Spiess BD. A comparison of a new ultrasound-based whole blood viscoelastic test (SEER sonorheometry) versus thromboelastography in cardiac surgery. *Anesth Analg*. 2016;123:1400–1407.
47. Shulman M, Myles P, Chan M, McIlroy D, Wallace S, Ponsford J. Measurement of disability-free survival after surgery. *Anesthesiology*. 2015;122:524–536.
48. Abu-Arafeh A, Jordan H, Drummond G. Reporting of method comparison studies: a review of advice, an assessment of current practice, and specific suggestions for future reports. *Br J Anaesth*. 2016;117:569–575.
49. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess*. 1998;2(i–iv):1–74.
50. Hunt TD, Ramanathan SA, Hannaford NA, et al. CareTrack Australia: assessing the appropriateness of adult healthcare: protocol for a retrospective medical record review. *BMJ Open*. 2012;2:e000665.
51. Hooper TD, Hibbert PD, Hannaford NA, et al. Surgical site infection-a population-based study in Australian adults measuring the compliance with and correct timing of appropriate antibiotic prophylaxis. *Anaesth Intensive Care*. 2015;43:461–467.

52. Wijeysundera DN. Predicting outcomes: is there utility in risk scores? *Can J Anaesth*. 2016;63:148–158.
53. Rathore SS, Weinfurt KP, Gross CP, Krumholz HM. Validity of a simple ST-elevation acute myocardial infarction risk index: are randomized trial prognostic estimates generalizable to elderly patients? *Circulation*. 2003;107:811–816.
54. Kalkman CJ, Visser K, Moen J, Bonsel GJ, Grobbee DE, Moons KG. Preoperative prediction of severe postoperative pain. *Pain*. 2003;105:415–423.
55. Sheth T, Chan M, Butler C, et al. Prognostic capabilities of coronary computed tomographic angiography before non-cardiac surgery: prospective cohort study. *BMJ*. 2015;350:h1907.
56. Goldman L, Weinberg M, Weisberg M, et al. A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain. *N Engl J Med*. 1982;307:588–596.
57. Canet J, Gallart L, Gomar C, et al. Prediction of postoperative pulmonary complications in a population-based surgical cohort. *Anesthesiology*. 2010;113:1338–1350.
58. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg*. 2013;217:833–842.e831–e833.
59. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med*. 2006;34:1297–1310.
60. American Statistical Association. What is a survey? www.amstat.org/sections/srms/whatsurvey.html. Accessed March 8 2018.
61. Jones D, Story D, Clavisi O, Jones R, Peyton P. An introductory guide to survey research in anaesthesia. *Anaesth Intensive Care*. 2006;34:245–253.
62. Tait AR, Voepel-Lewis T. Survey research: it's just a few questions, right? *Paediatr Anaesth*. 2015;25:656–662.
63. Klabunde CN, Willis GB, McLeod CC, et al. Improving the quality of surveys of physicians and medical groups: a research agenda. *Eval Health Prof*. 2012;35:477–506.
64. Story DA, Gin V, na Ranong V, Poustie S, Jones D. Inconsistent survey reporting in anesthesia journals. *Anesth Analg*. 2011;113:591–595.
65. Lohr KN, Steinwachs DM. Health services research: an evolving definition of the field. *Health Serv Res*. 2002;37:7–9.
66. Wijeysundera DN, Austin PC, Beattie WS, Hux JE, Laupacis A. Variation in the practice of preoperative medical consultation for major elective noncardiac surgery: a population-based study. *Anesthesiology*. 2012;116:25–34.
67. Moonesinghe SR, Wong DJN, Farmer L, Shawyer R, Myles PS, Harris SK. SNAP-2 EPICCS: the second Sprint National Anaesthesia Project-EPIdemiology of Critical Care after Surgery: protocol for an international observational cohort study. *BMJ Open*. 2017;7:e017690.
68. Lane-Fall MB, Beidas RS, Pascual JL, et al. Handoffs and transitions in critical care (HATRICC): protocol for a mixed methods study of operating room to intensive care unit handoffs. *BMC Surgery*. 2014;14:96.
69. Emond YE, Calsbeek H, Teerenstra S, et al. Improving the implementation of perioperative safety guidelines using a multifaceted intervention approach: protocol of the IMPROVE study, a stepped wedge cluster randomized trial. *Implement Sci*. 2015;10:3.
70. Harris AD, Bradham DD, Baumgarten M, Zuckerman IH, Fink JC, Perencevich EN. The use and interpretation of quasi-experimental studies in infectious diseases. *Clin Infect Dis*. 2004;38:1586–1591.
71. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919.
72. Paneth N, Fine P. The singular science of John Snow. *Lancet*. 2013;381:1267–1268.
73. Ho AMH, Phelan R, Mizubuti GB, et al. Bias in before-after studies: narrative overview for anesthesiologists. *Anesth Analg*. 2017.
74. Alvarez MP, Foley KE, Zebley DM, Fassler SA. Comprehensive enhanced recovery pathway significantly reduces postoperative length of stay and opioid usage in elective laparoscopic colectomy. *Surg Endosc*. 2015;29:2506–2511.
75. Streeter AJ, Lin NX, Crathorne L, et al. Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review. *J Clin Epidemiol*. 2017;87:23–34.
76. Wellek S, Blettner M. On the proper use of the crossover design in clinical trials: part 18 of a series on evaluation of scientific publications. *Dtsch Arztebl Int*. 2012;109:276–281.
77. Niesters M, Aarts L, Sarton E, Dahan A. Influence of ketamine and morphine on descending pain modulation in chronic pain patients: a randomized placebo-controlled cross-over proof-of-concept study. *Br J Anaesth*. 2013;110:1010–1016.
78. Colloca L, Miller FG. The nocebo effect and its relevance for clinical practice. *Psychosom Med*. 2011;73:598–603.
79. Sackett DL, Rosenberg WM. The need for evidence-based medicine. *J R Soc Med*. 1995;88:620–624.
80. Roberts C, Torgerson D. Randomisation methods in controlled trials. *BMJ*. 1998;317:1301.
81. Suresh K. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *J Hum Reprod Sci*. 2011;4:8–11.
82. Torgerson DJ, Roberts C. Understanding controlled trials. Randomisation methods: concealment. *BMJ*. 1999;319:375–376.
83. Wartolowska K, Beard D, Carr A. Blinding in trials of interventional procedures is possible and worthwhile. *F1000Res*. 2017;6:1663.
84. Candiotti KA, Kovac AL, Melson TI, Clerici G, Joo Gan T. A randomized, double-blind study to evaluate the efficacy and safety of three different doses of palonosetron versus placebo for preventing postoperative nausea and vomiting. *Anesth Analg*. 2008;107:445–451.
85. Finfer S, Bellomo R, Boyce N, French J, Myburgh J, Norton R. A comparison of albumin and saline for fluid resuscitation in the intensive care unit. *N Engl J Med*. 2004;350:2247–2256.
86. van Wijck AJ, Opstelten W, Moons KG, et al. The PINE study of epidural steroids and local anaesthetics to prevent postherpetic neuralgia: a randomised controlled trial. *Lancet*. 2006;367:219–224.
87. Mascha EJ. Equivalence and noninferiority testing in anesthesiology research. *Anesthesiology*. 2010;113:779–781.
88. Davidson AJ, Disma N, de Graaff JC, et al. Neurodevelopmental outcome at 2 years of age after general anaesthesia and awake-regional anaesthesia in infancy (GAS): an international multicentre, randomised controlled trial. *Lancet*. 2016;387:239–250.
89. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2009;361:1139–1151.
90. Sessler DI, Imrey PB. Clinical research methodology 3: randomized controlled trials. *Anesth Analg*. 2015;121:1052–1064.
91. Apfel C, Korttila K, Abdalla M, et al. A factorial trial of six interventions for the prevention of postoperative nausea and vomiting. *N Engl J Med*. 2004;350:2441–2451.
92. Myles PS, Smith JA, Forbes A, et al. Tranexamic acid in patients undergoing coronary-artery surgery. *N Engl J Med*. 2017;376:136–148.
93. Myles PS, Smith JA, Forbes A, et al. Continuing aspirin before coronary artery surgery. *N Engl J Med*. 2016;374:728–737.
94. Brown C, Lilford R. Evaluating service delivery interventions to enhance patient safety. *BMJ*. 2008;337:a2764.
95. de Smet AM, Kluytmans JA, Cooper BS, et al. Decontamination of the digestive tract and oropharynx in ICU patients. *N Engl J Med*. 2009;360:20–31.
96. Hillman K, Chen J, Cretikos M, et al. Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial. *Lancet*. 2005;365:2091–2097.
97. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*. 2015;350:h391.
98. Ellenberg SS. The stepped-wedge clinical trial: evaluation by rolling deployment. *JAMA*. 2018;319:607–608.
99. The Gambia Hepatitis Study Group. The gambia hepatitis intervention study. *Cancer Res*. 1987;47:5782–5787.
100. Moller AM, Myles PS. What makes a good systematic review and meta-analysis? *Br J Anaesth*. 2016;117:428–430.
101. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6:e1000097.
102. Pussegoda K, Turner L, Garritty C, et al. Systematic review adherence to methodological or reporting quality. *Syst Rev*. 2017;6:131.
103. Sivakumar H, Peyton PJ. Poor agreement in significant findings between meta-analyses and subsequent large randomized trials in perioperative medicine. *Br J Anaesth*. 2016;117:431–441.
104. The Cochrane Collaboration. RevMan5. London. http://www.cochrane.org.
105. The Cochrane Collaboration. Covidence. London. http://community.cochrane.org/tools/review-production-tools/covidence.

106. Debray TP, Moons KG, van Valkenhoef G, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Res Synth Methods*. 2015;6:293–309.
107. Thomas D, Radji S, Benedetti A. Systematic review of methods for individual patient data meta-analysis with binary outcomes. *BMC Med Res Methodol*. 2014;14:79.
108. Nevitt SJ, Marson AG, Davie B, Reynolds S, Williams L, Smith CT. Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. *BMJ*. 2017;357:j1390.
109. World Medical Association. WMA Declaration of Helsinki - Ethical principles for medical research involving human subjects. Ferney-Voltaire. www.wma.net/en/30publications/10policies/b3/index.html. Accessed March 8 2018.
110. Pildal J, Chan AW, Hrobjartsson A, Forfang E, Altman DG, Gotzsche PC. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. *BMJ*. 2005;330:1049.
111. Hrobjartsson A, Pildal J, Chan AW, Haahr MT, Altman DG, Gotzsche PC. Reporting on blinding in trial protocols and corresponding publications was often inadequate but rarely contradictory. *J Clin Epidemiol*. 2009;62:967–973.
112. World Health Organization. Recommended format for a research protocol. http://www.who.int/rpc/research_ethics/format_rp/en/. Accessed April 8 2018.
113. ICH Expert Working Group. *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline: Integrated addendum to ICH E6(R1): Guideline for Good Clinical Practice E6(R2) Step 2*; 2015. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R2__Addendum_Step2.pdf. Accessed April 8 2018.
114. National Institutes of Health and Food and Drug Agency. NIH and FDA release protocol template for phase 2 and 3 IND/IDE clinical trials. Washington. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-064.html. Accessed March 6 2018.
115. Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med*. 2013;158:200–207.
116. Chan AW, Tetzlaff JM, Gotzsche PC, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ*. 2013;346:e7586.
117. International Committee of Medical Journal. In: *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*. 2017. http://www.icmje.org/icmjerecommendations.pdf.
118. Rios LP, Ye C, Thabane L. Association between framing of the research question using the PICOT format and reporting quality of randomized controlled trials. *BMC Med Res Methodol*. 2010;10:11.
119. Haynes RB. Forming research questions. *J Clin Epidemiol*. 2006;59:881–886.
120. Nizamuddin SL, Nizamuddin J, Mueller A, Ramakrishna H, Shahul SS. Developing a hypothesis and statistical planning. *J Cardiothorac Vasc Anesth*. 2017;31:1878–1882.
121. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ*. 2015;350:h2147.
122. Oude Rengerink K, Kalkman S, Collier S, et al. Series: pragmatic trials and real world evidence: paper 3. Patient selection challenges and consequences. *J Clin Epidemiol*. 2017;89:173–180.
123. Sackett DL. Explanatory and pragmatic clinical trials: a primer and application to a recent asthma trial. *Pol Arch Med Wewn*. 2011;121:259–263.
124. National Institutes of Health. NIH policy and guidelines on the inclusion of women and minorities as subjects in clinical research – amended November 2017. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-18-014.html.
125. Geller SE, Koch A, Pellettieri B, Carnes M. Inclusion, analysis, and reporting of sex and race/ethnicity in clinical trials: have we made progress? *J Womens Health (Larchmt)*. 2011;20:315–320.
126. Spong CY, Bianchi DW. Improving public health requires inclusion of underrepresented populations in research. *JAMA*. 2018;319:337–338.
127. Hughson JA, Woodward-Kron R, Parker A, et al. A review of approaches to improve participation of culturally and linguistically diverse populations in clinical trials. *Trials*. 2016;17:263.
128. Reisner SL, Poteat T, Keatley J, et al. Global health burden and needs of transgender populations: a review. *Lancet*. 2016;388:412–436.
129. Zuidgeest MGP, Welsing PMJ, van Thiel G, et al. Series: pragmatic trials and real world evidence: paper 5. Usual care and real life comparators. *J Clin Epidemiol*. 2017;90:92–98.
130. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med*. 1987;317:141–145.
131. Welsing PM, Oude Rengerink K, Collier S, et al. Series: pragmatic trials and real world evidence: paper 6. Outcome measures in the real world. *J Clin Epidemiol*. 2017;90:99–107.
132. Kyte D, Duffy H, Fletcher B, et al. Systematic evaluation of the patient-reported outcome (PRO) content of clinical trial protocols. *PLoS One*. 2014;9:e110229.
133. Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: The SPIRIT-PRO Extension. *JAMA*. 2018;319:483–494.
134. Myles PS, Devereaux PJ. Pros and cons of composite endpoints in anesthesia trials. *Anesthesiology*. 2010;113:776–778.
135. Baker SG. Five criteria for using a surrogate endpoint to predict treatment effect based on data from multiple previous trials. *Stat Med*. 2018;37:507–518.
136. The CAST Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med*. 1989;321:406–412.
137. Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *JAMA*. 2002;288:358–362.
138. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med*. 2007;4:e297.
139. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Trials*. 2010;11:32.
140. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol*. 2010;63:e1–37.
141. Cook JA, Hislop J, Adewuyi TE, et al. Assessing methods to specify the target difference for a randomised controlled trial: DELTA (Difference ELicitation in TriAls) review. *Health Technol Assess*. 2014;18:1–175. v–vi.
142. Abdulatif M, Mukhtar A, Obayah G. Pitfalls in reporting sample size calculation in randomized controlled trials published in leading anaesthesia journals: a systematic review. *Br J Anaesth*. 2015;115:699–707.
143. Mascha EJ, Vetter TR. The statistical checklist and statistical review: two essential yet challenging deliverables. *Anesth Analg*. 2017;124:719–721.
144. ICH Expert Working Group. *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline: Statistical principles for Clinical trials (E9) Current Step 4 version*dated 5 February ; 1998. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf. Accessed April 7 2018.
145. Thomas L, Peterson ED. The value of statistical analysis plans in observational research: defining high-quality research from the start. *JAMA*. 2012;308:773–774.
146. Lang TA, Altman DG. Basic statistical reporting for articles published in biomedical journals: the "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines. *Int J Nurs Stud*. 2015;52:5–9.
147. Eisenach JC, Kheterpal S, Houle TT. Reporting of observational research in anesthesiology: the importance of the analysis plan. *Anesthesiology*. 2016;124:998–1000.
148. Thabane L, Wells G, Cook R, et al. Canadian-led capacity-building in biostatistics and methodology in cardiovascular and diabetes trials: the CANNeCTIN Biostatistics and Methodological Innovation Working Group. *Trials*. 2011;12:48.
149. Story D, Leslie K, French C. Feasibility and pilot studies. *Anaesth Intensive Care*. 2018.
150. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ*. 2016;355:i5239.
151. Kraemer HC, Mintz J, Noda A, Tinklenberg J, Yesavage JA. Caution regarding the use of pilot studies to guide power calculations for study proposals. *Arch Gen Psychiatry*. 2006;63:484–489.
152. Kirkegaard H, Soreide E, de Haas I, et al. Targeted temperature management for 48 vs 24 hours and neurologic outcome after out-of-hospital cardiac arrest: a randomized clinical trial. *JAMA*. 2017;318:341–350.

153. Bro-Jeppesen J, Hassager C, Wanscher M, et al. Targeted temperature management at 33 degrees C versus 36 degrees C and impact on systemic vascular resistance and myocardial function after out-of-hospital cardiac arrest: a sub-study of the target temperature management trial. *Circ Cardiovasc Interv*. 2014;7:663–672.
154. Frydland M, Kjaergaard J, Erlinge D, et al. Usefulness of serum b-type natriuretic peptide levels in comatose patients resuscitated from out-of-hospital cardiac arrest to predict outcome. *Am J Cardiol*. 2016;118:998–1005.
155. Gilje P, Koul S, Thomsen JH, et al. High-sensitivity troponin-T as a prognostic marker after out-of-hospital cardiac arrest - A targeted temperature management (TTM) trial substudy. *Resuscitation*. 2016;107:156–161.
156. Devereaux P, Sessler D, Leslie K, et al. Clonidine in patients having non-cardiac surgery. *N Engl J Med*. 2014;370:1504–1513.
157. Sessler DI, Meyhoff CS, Zimmerman NM, et al. Period-dependent associations between hypotension during and for four days after noncardiac surgery and a composite of myocardial infarction and death: a substudy of the POISE-2 trial. *Anesthesiology*. 2018;128:317–327.
158. Devereaux P, Sessler D, Mrkobrada M, et al. Aspirin in patients having non-cardiac surgery. *N Engl J Med*. 2014;370:1494–1503.
159. Eikelboom JW, Kearon C, Guyatt G, et al. Perioperative aspirin for prevention of venous thromboembolism: the perioperative ischemia evaluation-2 trial and a pooled analysis of the randomized trials. *Anesthesiology*. 2016;125:1121–1129.
160. Elze MC, Gregson J, Baber U, et al. Comparison of propensity score methods and covariate adjustment: evaluation in 4 cardiovascular studies. *J Am Coll Cardiol*. 2017;69:345–357.
161. Leslie K, Myles P, Devereaux P, et al. Neuraxial block, death and serious cardiovascular morbidity and mortality in the POISE Trial. *Br J Anaesth*. 2013;111:382–390.
162. Department of Health and Human Services. Office for human research protections. code of federal regulations title 45; public welfare, Department of Health and Human Services part 46: protection of human subjects. https://www.hhs.gov/ohrp/regulations-and-policy/regulations/index.html. Accessed April 7 2018.
163. Canadian Institutes of Health Research. Natural Sciences and Engineering Research Council of Canada; Social Sciences and Humanities Research Council of Canada. Tri-council policy statement: ethical conduct for research involving humans. http://www.pre.ethics.gc.ca/pdf/eng/tcps2-2014/TCPS_2_FINAL_Web.pdf. Accessed April 7 2018.
164. Health Research Authority. UK policy framework for health and social care research v3.3 07/11/17. London. https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/uk-policy-framework-health-social-care-research/. Accessed March 8 2018.
165. Australian Government. National Health and Medical Research Council; Australian Research Council. National statement on ethical conduct in human research 2007 (Updated May 2015). Canberra. https://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research. Accessed April 7 2018.
166. European Commission. Proposal for a regulation of the European Parliament and of the Council on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC. https://ec.europa.eu/health/sites/health/files/files/clinicaltrials/2012_07/proposal/2012_07_proposal_en.pdf. Accessed April 7 2018.
167. Australian Government: National Health and Medical Research Council. Ethical considerations in quality assurance and evaluation activities. Canberra. https://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/e111_ethical_considerations_in_quality_assurance_140326.pdf. Accessed April 7 2018.
168. Murphy GS, Szokol JW, Avram MJ, et al. Consent for anesthesia clinical trials on the day of surgery: patient attitudes and perceptions. *Anesthesiology*. 2016;124:1246–1255.
169. Chludzinski A, Irani C, Mascha EJ, Kurz A, Devereaux PJ, Sessler DI. Protocol understanding and anxiety in perioperative clinical trial patients approached for consent on the day of surgery. *Mayo Clin Proc*. 2013;88:446–454.
170. Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol*. 1986;4:1529–1541.
171. Zarin DA, Tse T, Williams RJ, Rajakannan T. Update on trial registration 11 years after the ICMJE policy was established. *N Engl J Med*. 2017;376:383–391.
172. US Department of Health and Human Services. Food and Drug Administration, Center for Drug Evaluations and Research, Center for Biologics Evaluation and Research. Guidance for Industry and Investigators: Safety Reporting Requirements for INDs and BA/BE Studies. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM227351.pdf. Accessed April 7 2018.
173. European Federation of Pharmaceutical Industries and Associations (EFPIA), the International Federation of Pharmaceutical Manufacturers and Associations (IFPMA), the Japanese Pharmaceutical Manufacturers Association (JPMA) and the Pharmaceutical Research and Manufacturers of America (PhRMA). Joint position on the disclosure of clinical trial information via clinical trial registries and databases. Updated November 10, 2009 (with minor revisions as of October 30, 2017). https://www.ifpma.org/wp-content/uploads/2010/11/Joint-Position-on-Disclosure-of-CT-Info-via-CT-Registries-Revised-Oct2017-vF.pdf. Accessed April 8 2018.
174. European Commission. Commission Guideline — Guidance on posting and publication of result-related information on clinical trials in relation to the implementation of Article 57(2) of Regulation (EC) No 726/2004 and Article 41(2) of Regulation (EC) No 1901/2006. https://ec.europa.eu/health//sites/health/files/files/eudralex/vol-10/2012_302-03/2012_302-03_en.pdf. Accessed April 8 2018.
175. Jones PM, Chow JTY, Arango MF, et al. Comparison of registered and reported outcomes in randomized clinical trials published in anesthesiology journals. *Anesth Analg*. 2017;125:1292–1300.
176. Nizamuddin J, Nizamuddin SL, Shahul S. Trial registration and outcome reporting: a bait and switch? *Anesth Analg*. 2017;125:1098–1099.
177. Loder E, Groves T, Macauley D. Registration of observational studies. *BMJ*. 2010;340:c950.
178. Sorensen HT, Rothman KJ. The prognosis for research. *BMJ*. 2010;340:c703.
179. European Medicines Agency. Clinical trials in human medicines. http://www.ema.europa.eu/ema/index.jsp?curl=pages/special_topics/general/general_content_000489.jsp&mid=WC0b01ac058060676f. Accessed March 6 2018.
180. ICH Expert Working Group. *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline: Development and manufacture of drug substances (Chemical entities and biotechnological/biological entities) Q11. Step 4 version dated*; 2012. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q11/Q11_Step_4.pdf. Accessed April 8 2018.
181. Absalom AR, Glen JI, Zwart GJ, Schnider TW, Struys MM. Target-controlled infusion: a mature technology. *Anesth Analg*. 2016;122:70–78.
182. Institute of Medicine (IOM). Sharing clinical trial data: maximizing benefits, minimizing risk. Washington. http://nationalacademies.org/hmd/reports/2015/sharing-clinical-trial-data.aspx. Accessed March 6 2018.
183. Pharmaceutical Research and Manufacturers of America, European Federation of Pharmaceutical Industries and Associations. Principles for responsible clinical trial data sharing. http://phrma-docs.phrma.org/sites/default/files/pdf/PhRMAPrinciplesForResponsibleClinicalTrialDataSharing.pdf. Accessed April 8 2018.
184. Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials. *BMJ*. 2017;357:j2372.
185. Ohmann C, Banzi R, Canham S, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open*. 2017;7:e018647.
186. Taichman DB, Backus J, Baethge C, et al. Sharing clinical trial data: a proposal from the International Committee of Medical Journal. *N Engl J Med*. 2016;374:384–386.
187. Tudur Smith C, Nevitt S, Appelbe D, et al. Resource implications of preparing individual participant data from a clinical trial to share with external researchers. *Trials*. 2017;18:319.
188. Herbert DL, Coveney J, Clarke P, Graves N, Barnett AG. The impact of funding deadlines on personal workloads, stress and family relationships: a qualitative study of Australian researchers. *BMJ Open*. 2014;4:e004462.
189. Herbert DL, Barnett AG, Clarke P, Graves N. On the time spent preparing grant proposals: an observational study of Australian researchers. *BMJ Open*. 2013;3.
190. Meinecke AK, Welsing P, Kafatos G, et al. Series: pragmatic trials and real world evidence: paper 8. Data collection and management. *J Clin Epidemiol*. 2017;91:13–22.
191. Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. *Perspect Clin Res*. 2014;5:159–166.
192. Morgan DJ, Ho KM. The anaesthetic assessment, management and risk factors of bariatric surgical patients requiring postoperative intensive care support: a state-wide, five-year cohort study. *Anaesth Intensive Care*. 2016;44:237–244.

193. National Institutes of Health. Good clinical practice training. https://grants.nih.gov/policy/clinical-trials/good-clinical-training.htm. Accessed April 8 2018.
194. Pearse RM, Harrison DA, MacDonald N, et al. Effect of a perioperative, cardiac output-guided hemodynamic therapy algorithm on outcomes following major gastrointestinal surgery: a randomized clinical trial and systematic review. *JAMA*. 2014;311:2181–2190.
195. Treweek S, Pitkethly M, Cook J, et al. Strategies to improve recruitment to randomised trials. *Cochrane Database Syst Rev*. 2018;2:MR000013.
196. Brueton VC, Tierney J, Stenning S, et al. Strategies to improve retention in randomised trials. *Cochrane Database Syst Rev*. 2013;(12):MR000032.
197. ICH Expert Working Group. *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use*. ICH Harmonised Tripartite Guideline: clinical safety data management: definitions and standards for expedited reporting E2A. Step 4 version dated 27 October; 1994. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2A/Step4/E2A_Guideline.pdf. Accessed April 8 2018.
198. ICH Expert Working Group. *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use*. ICH Harmonised Tripartite Guideline: development safety update report E2F. Step 4 version dated 17 August ; 2010. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2F/Step4/E2F_Step_4.pdf. Accessed April 8 2018.
199. Reith C, Landray M, Devereaux PJ, et al. Randomized clinical trials--removing unnecessary obstacles. *N Engl J Med*. 2013;369:1061–1065.
200. Wallace S, Myles PS, Zeps N, Zalcberg JR. Serious adverse event reporting in investigator-initiated clinical trials. *Med J Aust*. 2016;204:231–233.
201. European Medicines Agency, Heads of Medicines Agencies. Guideline on good pharmacovigilance practices (GVP). Module VI Collection, management and submission of reports of suspected adverse reactions to medicinal products (Rev 2). http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2017/08/WC500232767.pdf. Accessed April 8 2018.
202. Pogue JM, Devereaux PJ, Thorlund K, Yusuf S. Central statistical monitoring: detecting fraud in clinical trials. *Clin Trials*. 2013;10:225–235.
203. Fleming TR, DeMets DL, Roe MT, et al. Data monitoring committees: promoting best practices to address emerging challenges. *Clin Trials*. 2017;14:115–123.
204. EQUATOR Network. Enhancing the quality and transparency of health research. http://www.equator-network.org/. Accessed April 8 2018.
205. Dechartres A, Trinquart L, Atal I, et al. Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: research on research study. *BMJ*. 2017;357:j2490.
206. Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev*. 2012;11:MR000030.
207. Blanco D, Kirkham JJ, Altman DG, Moher D, Boutron I, Cobo E. Interventions to improve adherence to reporting guidelines in health research: a scoping review protocol. *BMJ Open*. 2017;7:e017551.
208. Fontanarosa P, Bauchner H, Flanagin A. Authorship and team science. *JAMA*. 2017;318:2433–2437.
209. Bierer BE, Crosas M, Pierce HH. Data authorship as an incentive to data sharing. *N Engl J Med*. 2017;377:402.
210. US National Library of Medicine. Number of authors per MEDLINE/PubMed citation. http://www.nlm.nih.gov/bsd/authors1.html. Accessed April 8 2018.
211. Committee on Publication Ethics (COPE). Core practices. Washington. https://publicationethics.org/core-practices. Accessed March 6 2018.
212. Shen C, Bjork BC. Predatory open access: a longitudinal study of article volumes and market characteristics. *BMC Med*. 2015;13:230.
213. Moher D, Shamseer L, Cobey KD, et al. Stop this waste of people, animals and money. *Nature*. 2017;549:23–25.
214. Directory of Open Access Journals. 2018;8.
215. Korenman SG. Research integrity. https://ori.hhs.gov/education/products/ucla/chapter1/page02.htm.
216. Office of Science and Technology Policy. Federal research misconduct policy. *Federal Register, December 6, 2000*. 2000;65:76260–76264.
217. No Authors Listed. A consensus statement on research misconduct in the UK. *BMJ*. 2012;344:e1111.
218. Haug CJ. Peer-review fraud - hacking the scientific publication process. *N Engl J Med*. 2015;373:2393–2395.
219. Marusic A, Wager E, Utrobicic A, Rothstein HR, Sambunjak D. Interventions to prevent misconduct and promote integrity in research and publication. *Cochrane Database Syst Rev*. 2016;4:MR000038.
220. Shafer SL. Plagiarism is ubiquitous. *Anesth Analg*. 2016;122:1776–1780.
221. Moylan EC, Kowalczuk MK. Why articles are retracted: a retrospective cross-sectional study of retraction notices at BioMed Central. *BMJ Open*. 2016;6:e012047.
222. Horsley T, Lockyer J, Cogo E, Zeiter J, Bursey F, Campbell C. National programmes for validating physician competence and fitness for practice: a scoping review. *BMJ Open*. 2016;6:e010368.

90 Interpreting the Medical Literature

ELIZABETH L. WHITLOCK and CATHERINE L. CHEN

KEY POINTS

- The medical literature can be accessed in many ways, ranging from the primary literature as indexed in sources like PubMed, to sources catering to medical professionals, and via the lay press and social media as well.
- The quality of medical literature varies widely, as does the interpretation or distillation into secondary sources of information.
- The traditional "evidence pyramid" no longer effectively reflects the diversity and impact on clinical care of the rapidly evolving body of medical knowledge.
- Understanding the design of a research study is crucial to understanding the strength and implications of its conclusions.
- Different research study designs have assessment tools, maintained by the EQUATOR Network, to assist in understanding the quality of the study.
- Statistical analysis is commonplace in original research manuscripts, but the interpretation of statistical results, particularly the probability or "*P*-value," is widely misunderstood and misused.
- Although the hallmark of the primary medical literature is a rigorous peer review process, there are serious pitfalls that may be difficult to detect, including predatory journals and reviewer misconduct.

Introduction

Reading and interpreting the medical literature is a critical skill for any practicing anesthesiologist. The numerous scientific advances made over the last few decades make it imperative for anesthesiologists to understand how to read, interpret, and apply the medical literature to the clinical scenarios that arise in daily practice. **Critical appraisal,** a concept that first arose in the early 1980s,[1] is "the process of carefully and systematically examining research to judge its trustworthiness, and its value and relevance in a particular context."[2] The idea that clinicians "should practice medicine in a way that combines research evidence with clinical skills and patient values and preferences" led to the creation of a new approach to caring for patients called "evidence-based medicine," a term first coined in the early 1990s by Gordon Guyatt at McMaster University.[3]

In response to this new approach to patient care, where the most recent published evidence could be used to support clinical decisions in lieu of the usual approach of deferring to a physician's seniority, intuition, or prior experiences with similar patients, the *Journal of the American Medical Association* published a series of articles entitled "Users' Guides to the Medical Literature" in the early 1990s that has since been formatted into a textbook by the same name.[4,5] This text is an excellent resource for those who are interested in effectively incorporating the systematic evaluation of the medical evidence into their daily clinical practice. We do not purport to write as complete a chapter on interpreting the medical literature as the existing resources that are already available to the lay reader.

This chapter is, however, intended to provide some basic information about how medical evidence is created and published, some useful tools for evaluating the medical literature, and to highlight some of the pitfalls to avoid when sorting through the litany of information available to both physicians and to the lay public. The previous chapter provided an introduction to the different study designs used in clinical research and their relative strengths and weaknesses. This chapter will aim to put the key points raised in the previous chapter within the context of how research can be interpreted and used by a clinical anesthesiologist.

INTENT OF THE CHAPTER

The goals of this chapter are: (1) To provide a brief overview of how a research manuscript is treated from the moment of submission to its publication date; (2) to provide a practical guide for accessing, interpreting the quality of, and implementing the knowledge gained from reading the medical literature; and (3) to identify and avoid the pitfalls of indiscriminate use or the misuse of the published evidence.

The Publication Process

TYPES OF JOURNALS

With the advent of the Internet, researchers and consumers of research have an overwhelming array of journals to consider for submitting and reading original work. A journal's focus and target audience can vary. Some journals are

known as **general medical journals** because they include work from many different fields within medicine. Examples include *The New England Journal of Medicine (NEJM)*,[6] *Journal of the American Medical Association (JAMA)*,[7] *The BMJ (formerly known as the British Medical Journal)*,[8] and *The Lancet*.[9] Most medical journals, however, focus on a particular medical specialty, or they may highlight particular topics such as research methods or health policy.

Journals can be accessed as printed and bound periodicals requiring an active print subscription, through a designated website offering a selection of free articles or articles requiring payment to view, or often by using a combination of the two delivery methods. The quality and reputation of journals can vary greatly—some require little to no peer review and an upfront payment in exchange for quick online publication of a manuscript of questionable quality (see *Predatory Journals*, later). Others maintain extremely high standards for manuscript acceptance and editorial review. Publication in a medical journal with an established history as a traditional print journal, even while it maintains a distinct web presence, is usually an indicator that a published manuscript has been subjected to peer review and represents a worthwhile contribution to the medical literature.

A number of professional societies in anesthesiology publish journals of peer-reviewed literature relevant to the field and include some of the most reputable journals within the field, such as the American Society of Anesthesiologists *(Anesthesiology)*,[10] the Royal College of Anaesthetists/The College of Anaesthetists of Ireland/the Hong Kong College of Anaesthesiologists *(British Journal of Anaesthesia)*,[11] and the International Anesthesia Research Society *(Anesthesia & Analgesia)*.[12]

TYPES OF JOURNAL ARTICLES

Although each journal's priorities for content may vary, for the most part, journals publish similar types of articles. These article types can be grouped into the following broad categories: original research, review articles, brief reports or letters, case reports, and editorials.

Original research is the most familiar and most common type of article that gets published. Original research usually consists of a scientific manuscript that reports the full results from a research study and can represent any of the study designs that were described in Chapter 89.

Review articles summarize the existing scientific research on a given topic and are a good way for readers to quickly familiarize themselves with the current evidence within an area of research. Review articles are comprehensive, typically written by experts in the field, and are often solicited by the editors of a journal. The authors will usually frame published research within the context of other contemporaneous works and the current and future directions of the research topic being reviewed.

Brief reports and letters provide concise research reports that address a timely issue or spur further research if published before a full original research manuscript has been submitted. Letters are also an opportunity for readers to submit arguments extending or rebutting articles that were previously published in that journal.

Case studies allow researchers or clinicians to share specific examples of unusual or unexpected clinical findings in a single patient that may be instructive to a broader readership. A case series is similar to a case study but describes similar clinical phenomena across multiple patients.

Editorials are essays that encompass the opinion of the author about a subject, usually of topical relevance, or highlight the important scientific contributions of an original research article that was published in the same journal issue. Similar to review articles, editorials are usually solicited by the journal editor, written by experts in the field, and provide important context with which to frame the original research article.

THE PEER REVIEW PROCESS

The peer review process is an important component of publishing original research. Most respected journals, whether web-based or traditional print journals, will have a robust peer review process in place. Once an original research manuscript has been submitted by the authors, the journal editors will usually make a rapid decision about whether the study topic is appropriate for their readership. Authors will either receive notification that their submission was rejected without review, or that the manuscript has been sent out for review to at least two other experts in the field. Reviewers are asked to rate various aspects of the manuscript, including its **readability, novelty, methods, validity of results**, and **potential impact** on the field. They will often provide constructive feedback to the authors on ways the study can be substantially improved. Reviewers will then make a recommendation to the journal editors whether the study should be accepted, revised and resubmitted, or rejected. The editors will consider the reviewers' comments in making the final decision about the manuscript's disposition.

Although different journals may have different terminologies for these final decisions, in general, they fall into the following categories: accept as written, conditional acceptance (i.e., accept with minor revisions or accept with major revisions), revise and resubmit, or reject. **Acceptance** of a manuscript without any changes is extremely rare. **Conditional acceptance**, though frequently considered to be a positive outcome, is not a guarantee of acceptance. The editors still retain the right to reject the manuscript unless the manuscript's authors have satisfactorily addressed any concerns that are deemed important enough to be raised in the outcome letter. More common is a decision to **revise and resubmit**. The editor may require substantial changes to the original manuscript and usually requests a new manuscript submission that clearly highlights within the text what changes have been made in response to the outcome letter. Most of the time, the authors are also required to submit an accompanying document that responds to each of the points that may have been raised by the editor and reviewers during the peer review process. Finally, the editors can still decide to **reject** a manuscript after sending it out for review.

Once a manuscript has been through the peer review process and deemed acceptable for publication, the journal will then format the entire manuscript to fit the journal's style. This usually includes recreating or reformatting the tables

and figures that were submitted with the manuscript, as well as detailed copyediting for grammar, punctuation, and clarity. **Proofs** of the formatted article, which show how the article will actually appear when it is finally printed, are sent to the author for final approval. Simultaneously, the journal will select the print issue that will feature the accepted article and decide whether the manuscript would benefit from an accompanying editorial. Although print journals usually require a lead time of several months after an original article has been accepted before it is actually published, a journal will often set an earlier date for online publication, typically referred to as "e-pub ahead of print." Expediting the dissemination of original research by publishing electronically on a journal's website allows readers to obtain the earliest access to interesting and timely research.

Accessing the Medical Literature

PRIMARY LITERATURE

In the past, readers needed to have a journal subscription or access to a medical library to read published articles. In the academic setting, it was very common for attending physicians to photocopy important articles and distribute them to their trainees. As in so many other arenas, access to research articles has been transformed with the arrival of the Internet, and most articles are now easily accessible online, either directly through the journal's website, or through indexed search engines. **Primary literature** refers to these original research articles that are authored by the researchers who performed the study and are published in peer-reviewed journals.[13]

Most readers are probably familiar with **PubMed**,[14] a free resource maintained by the National Center for Biotechnology Information[15] at the US National Library of Medicine,[16] which is located at the National Institutes of Health.[17] PubMed provides access to MEDLINE,[18] an online database that contains more than 28 million references to journal articles in life sciences, with a concentration on biomedicine. The database includes medical literature published from 1966 to the present day, with citations from more than 5200 journals in approximately 40 languages worldwide, with new articles being added daily. Search results include a list of article citations with links to the electronic full-text article if available. Although PubMed is most frequently used to access the primary literature, this resource is useful for accessing the secondary literature as well, which is described in the following section. Searching PubMed is free of charge, and many articles are available for viewing without a subscription or an academic institutional affiliation, including through the archival of full-text articles in PubMed Central. PubMed provides various mechanisms to access articles that are not freely available on the Internet, although fees may apply.

SECONDARY LITERATURE

In addition to primary sources, physicians may rely on other resources to remain informed about recent advances in their field. The term **secondary literature** is used to refer to written summaries of the primary literature that help synthesize or evaluate primary sources for the purpose of dissemination and incorporation of evidence-based medicine into practice.[13] These articles summarize the primary literature to varying degrees and with variable quality, depending on their purpose. **Systematic reviews** and **metaanalyses** are, in and of themselves, valued as high-quality research and important contributions to the medical literature. For example, the Cochrane Database of Systematic Reviews[19] is a well-known and highly regarded resource for systematic reviews in health care. Readers should be aware, however, that **narrative reviews** differ from systematic reviews in that they are not obliged to provide unbiased information reflecting the totality of available knowledge. Narrative reviews can be useful, and an efficient source of information particularly when written by authors who might be expected to have an expert command of the available literature, but fundamentally differ from systematic reviews because they are not as rigorous or comprehensive in their approach to manuscript selection, which must be considered in their interpretation. Types of reviews and metaanalyses are further discussed in Chapter 89.

Clinical practice guidelines fall into the category of secondary literature as well. These are typically written by professional groups or government agencies to help guide clinician decision making and will often indicate the level of evidence supporting the practice recommendations that are being made. Other more filtered—yet still evidence-based—distillations of the existing research exist to facilitate clinical decision making by practicing clinicians at the bedside. Examples of the latter include websites like UptoDate[20] and WebMD.[21]

TRADITIONAL AND SOCIAL MEDIA

Finally, the medical literature can be indirectly accessed via both traditional and social media. Traditional media may include press releases issued by the author's institution or coordinated by the print journal to maximize the impact and newsworthiness of an important scientific advance. These press releases can often lead to articles in leading newspapers or newsmagazines that refer to the original article if they are considered of interest to the general public. However, traditional media can also inadvertently or deliberately present an inaccurate or sensationalized version of the article's conclusions, potentially not tempered by the study's limitations. Media reports of peer-reviewed articles are not necessarily written in collaboration with the authors of the article, and even though quotations may be excerpted from press releases and other sources, conclusions can be misrepresented. If a layperson's interpretation of a scientific finding is to be used for clinical practice, the validity of the conclusions presented must be verified against the original article itself.

Many scientists and clinicians are now keeping current with the latest evidence through social media sites such as Twitter[22] and Facebook,[23] or individual blogs, where leading scientists, researchers, and clinicians can link to original articles and provide their own spin or commentary on the relative merits or shortcomings of the most recently published research. Other less reliable, but easily accessible, resources include crowd-sourced websites such as Wikipedia.[24] However, the quality of the information presented on individual blogs or crowd-sourced sites depends

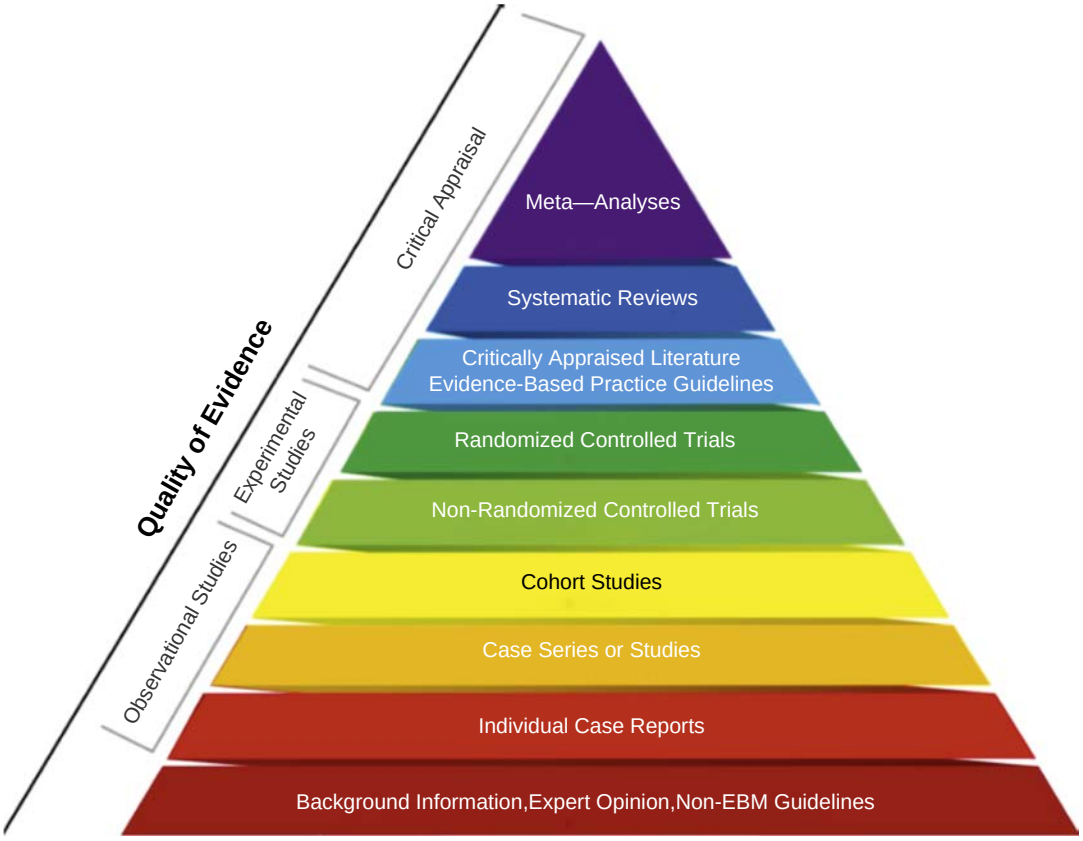

Fig. 90.1 The Evidence Pyramid.

significantly on the credentials of the individuals contributing to the website. Although the democratization of the medical literature has accelerated the dissemination of new research to both scientific audiences and to the public, it is still important for individual clinicians to understand how to approach the literature independently to separate the propaganda surrounding the latest research from the actual strengths and weaknesses of the study itself.

Assessing the Methodology of a Study

Understanding how and why a study was done is essential to understanding how its findings fit in with the progress of medical science. The quality of clinical research is highly dependent on many design choices that were made long before recruitment of the first patient, or collection of the first data record, and a familiarity with clinical research methodology will help all medical professionals critically evaluate the applicability of published findings to their own practice decisions.

As a first pass, it is easiest to separate research designs into two bins: **observational** and **interventional**. Within those designations are subcategories and variants with important implications for study quality. Furthermore, some studies such as metaanalyses incorporate aspects of both. Contrary to older notions of research interpretation, a randomized trial does not always produce better evidence than an observational (cohort) study, and a metaanalysis—far from being the "pinnacle"—is only as good as the evidence on which it rests.

THE "EVIDENCE PYRAMID" AND ITS EVOLUTION

Historically, those interested in evaluating study quality have been referred to an "evidence pyramid," where each step up the pyramid represents a step closer to truth (or quality, or best evidence) (Fig. 90.1).[25]

This pyramid produces a striking visual representation emphasizing the weaknesses—and relative abundance—of expert opinion and observational case studies, the importance of randomized controlled experimental trials, and the primacy of more recently developed summative methods, including systematic reviews and metaanalyses, which comprise the apex of evidentiary quality. But this is overly simplistic in today's world of complex patients with multiple intersecting conditions: the pyramid, with its unlabeled y-axis that may be thought of as "risk of bias" or "internal validity," rather than a progress toward "truth," is due for modification.[26]

Two attempts to refine this pyramid are in Figs. 90.2A and B. Fig. 90.2A highlights the variability in quality within a single type of study design, and the dependence of summative methods like systematic reviews upon the quality of existing evidence,[26] whereas Fig. 90.2B avoids the appearance of a hierarchy altogether, emphasizing that data derived from different methods is necessary to provide a strong foundation for scientific knowledge.[27] Another graphical reconceptualization of the way scientific evidence is generated was proposed by Walach and colleagues in 2006 (see Fig. 90.2C), and recently updated toward a "matrix" concept: their "Circle of Methods" provides a more granular categorization that differentiates between efficacy

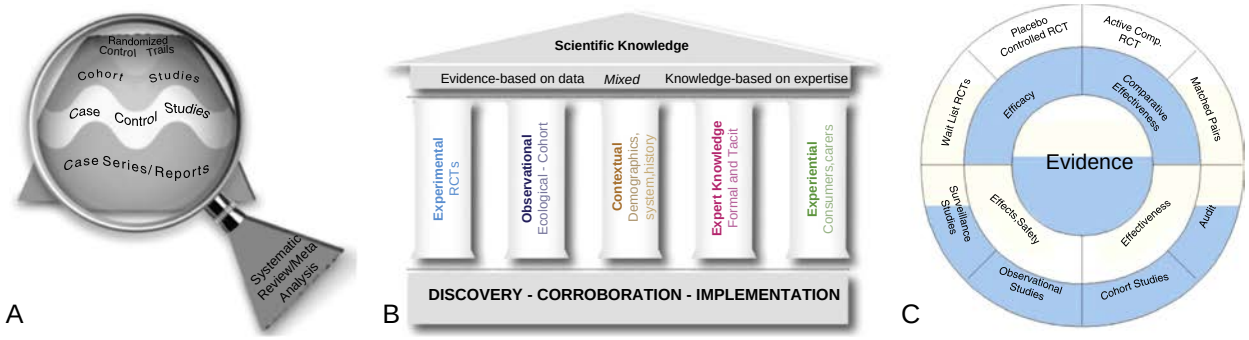

Fig. 90.2 Proposed refinements of the "Evidence Pyramid," reflecting the lack of consensus in the scientific community about how to visually encapsulate the relationships among methods of generating evidence. (A) A refinement proposed by Murad and colleagues. (B) The Greek Temple model. (C) Circle of Methods. ([A], Redrawn from Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. *Evid Based Med*. 2016;21(4):125–127. [B] Redrawn from Salvador-Carulla L, Lukersmith S, Sullivan W. From the EBM pyramid to the Greek temple: a new conceptual approach to guidelines as implementation tools in mental health. *Epidemiol Psychiatr Sci*. 2017;26(2):105–114. [C], Redrawn from Tugwell P, Knottnerus JA. Is the evidence pyramid now dead? *J Clin Epidemiol*. 2015;68(11):1247–1250.)

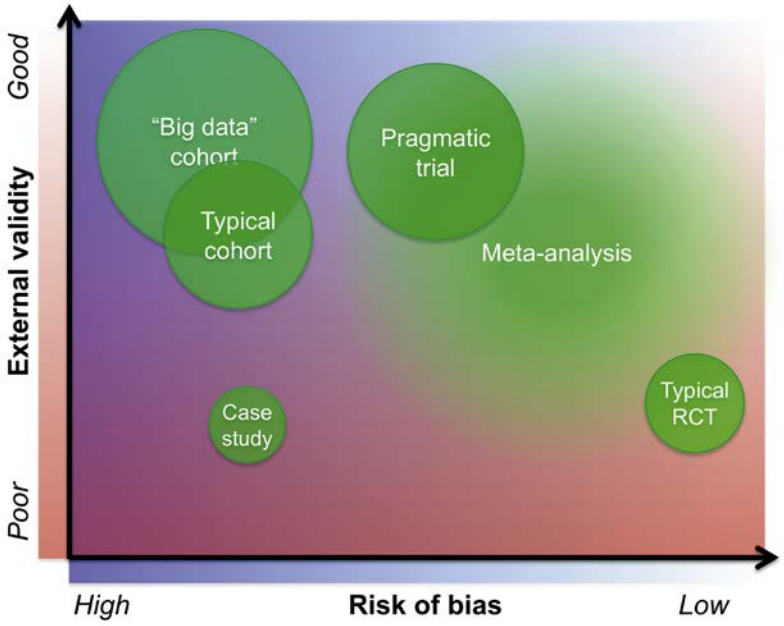

Fig. 90.3 The relationships between sample size (often implying higher costs), external validity, and risk of bias produces trade-offs in study design. Here, bubble size is proportional to the typical number of participants in a particular type of study (larger bubble indicates more participants).

and effectiveness, a crucial concept in the evolution of medical knowledge toward changing care for broad populations of patients.[28]

Perhaps the issue is that no single graphic can capture the unique strengths and limitations of the common experimental designs upon which the advancement of medical science relies. The perfect study would have no bias and a high level of external validity, reflecting some scientific truth that is universally applicable. This is probably impossible to do. Instead, when choosing among different clinical research designs, researchers are faced with many trade-offs between internal and external validity, feasibility (including costs), and risk of bias (Fig. 90.3).

BASIC RESEARCH DESIGNS

Trade-offs between observational and interventional designs are discussed more fully in Chapter 89. Briefly, traditional **observational studies** include cohort studies, case-control studies, and cross-sectional studies, which are unified by their non-interventional nature. This results in risk of bias, which can be mitigated (but not eliminated) by sensible experimental design and analytic choices. **Interventional trials** are the most used method of generating data to demonstrate causal relationships. While observational studies are subject to confounding by observed or unobserved variables, interventional trials take variable advantage of features meant to minimize bias: use of **controls, randomization**, and **blinding**. Understanding the implications of experimental design choices in these three areas, which are discussed in Chapter 89, underpins the assessment of trial quality.

As mentioned earlier in this chapter, summative studies like systematic reviews and metaanalyses offer another method of objectively summarizing available evidence and are discussed further in Chapter 89. Just as a summary of medical evidence presented on a layperson's news website must be approached with caution, a summary of evidence

via metaanalysis or systematic review nonetheless glosses over what may be important differences in the quality and methodology of the underlying studies, and is highly reliant on the skill and thoroughness of its authors.[29] Far from being an easy way to understand "scientific truth" in the literature, summative methods of generating evidence hold as many pitfalls as their underlying trials. However, in skilled hands and with the right underlying materials, systematic reviews and metaanalyses can provide important evidence, which could not feasibly be generated any other way.

BIG DATA AND PRAGMATIC CLINICAL TRIALS

"Big data" studies have a unique and evolving place in the continuum of evidence. With the advent and widespread adoption of electronic medical records, enormous quantities of data are recorded every day during the routine provision of clinical care. This has led to two major evolutions in clinical research: the large **secondary data cohort study** and its interventional correlate, the **pragmatic clinical trial**.

Traditionally, data collection decisions for prospective cohort studies were based on the research query itself: for example, a study of lung function over time would obtain annual formal pulmonary function tests. All the anticipated necessary variables would be collected—height, weight, pulmonary function test results, chest x-rays, medication lists, exercise habits, dust mite exposure, and a detailed tobacco use history. From these variables, a (potentially extensive) list of confounding variables would be selected and, in mathematical modeling, accounted for, thus producing an estimate of the "independent effect" of the primary predictor, for example, air pollution exposure on the outcome of lung function.

Collecting extensive data is expensive, though, and requires time and resources. What if dust mite exposure and exercise habits are not hypothesized to be particularly important—could those be eliminated? What if, for no cost and only the time of an analyst to extract the data from an electronic medical record, we could use current tobacco use as a surrogate for a thorough tobacco use history, instead of hiring a clinical research nurse to collect the information? And if we felt that the height, weight, pulmonary function tests, and smoking history collected in the course of normal clinical care would be of sufficient quality, perhaps we could estimate air pollution exposure levels based on proximity to major roads—and, by eliminating the need to directly collect any specific data on our population, expand our population target to all those with valid values of these variables in our electronic medical records system? Thus, the "big data" study is born.

A more generic term for "big data" is "**secondary data**": data collected for purposes other than the study of a researcher's intended question. There are important fundamental differences between a traditional cohort study and the approach described here. Inherent in the compromise is a move from detailed and specific characterization of each participant to a more diffuse idea of individual participant characteristics, and a move from a small (and likely specific) population, compromising external validity, to one that is large and generalizable. Some phenomena simply cannot be studied any other way because they involve a subtle effect or rare outcome requiring large populations to show an association that exceeds our currently acceptable margins of uncertainty in strength—hence the term "big data." Some phenomena are particularly suited to study using this approach, because (acceptably) high-fidelity data are collected in the course of routine clinical care. One such example might be studies of association between postoperative respiratory adverse events and intraoperative mechanical ventilation parameters. Studies on the epidemiology of medical care, or its financial costs, are highly dependent on the availability and quality of secondary data.

There is a growing appreciation that the data collection infrastructure that supports the provision of daily medical care can also be used to support large-scale **pragmatic clinical trials**. Although the philosophy of a pragmatic trial was described before the widespread implementation of electronic medical records,[30] these studies have been greatly facilitated by the increase in availability of secondary data. In contrast to the typical randomized controlled trial, a pragmatic trial seeks to avoid restrictive inclusion criteria and highly protocolized care, more closely recapitulating the way medical care is provided to individual patients. They are typically categorized as **effectiveness**, rather than **efficacy**, trials. This means that findings may have excellent **external validity** (applicability to a broad population of patients and settings). But since the lack of protocolization may introduce random or nonrandom variability in care, these trials are typically very large: thousands or tens of thousands of participants. Pragmatic trials may also be more vulnerable to risk of bias than traditional randomized trials, since it may not be practical or possible to blind clinicians and/or participants, in addition to challenges introduced by using secondary data (described earlier).

One recent pragmatic trial example relevant to anesthesiology is the SMART trial, published in *NEJM*, which randomized 15,802 critically ill patients to receive normal saline (0.9% sodium chloride) or a balanced salt solution.[31] The pragmatic features in this trial included: a cluster-randomized design, where the ICU in which care was provided determined which solution the patient received (instead of individual randomization, which introduces additional protocol adherence challenges); use of the electronic medical record to prompt ordering providers to consider relative contraindications and, if none were present, follow the protocol (rather than having study interventions delivered only by study personnel); and outcome and adjustment variable collection by the electronic medical record (i.e., use of secondary data). Importantly, both interventions—the use of saline versus a balanced solution when intravenous crystalloid solution is required—are standard of care, and the requirement for informed participant consent was, in this trial, waived. The target sample size was achieved in fewer than 2 years of enrollment, with less than 1 year between completion of enrollment and publication. Concurrently, a second complementary trial—of saline versus balanced crystalloid in noncritically ill adults receiving intravenous fluids in the emergency department (SALT-ED trial)—was also run, enrolling over 13,000 patients with a similar randomization scheme and data collection methodology.[32] These two trials substantially contribute to a long-running debate in medicine, and are excellent examples of pragmatic clinical trials.

Evaluating the quality of a study based on data collected for other purposes is complex, challenging, and beyond the

scope of this chapter. Fundamental differences in bias, generalizability, and even primary findings can be wrought by small and potentially undetectable design or analytic choices, such as geographically or socioeconomically limited populations; systematic problems in data quality resulting from misaligned incentives; inclusion or exclusion (by choice, or because of unavailability) of key confounding variables; presence of missing data and its treatment; statistical coding errors; and so on. Nevertheless, secondary data research has supported preliminary hypotheses that have launched innumerable research inquiries generating measurable improvements in human health, has allowed us to define the scope and costs of the medical care provided today, offers the opportunity to study research questions that would otherwise be completely inaccessible because of cost or ethical barriers, and holds enormous promise in the burgeoning era of "personalized medicine."

Tools to Ensure Study Quality

RECOGNITION OF THE NEED FOR STANDARDS IN STUDY REPORTING

Research papers are a massive simplification of months or years of work with thousands or maybe millions of choices encapsulated in 3000 words or fewer. Complete disclosure of all details of a work is neither practical nor desirable; the current standard is generally thought of as providing sufficient detail so that another investigator could reproduce the study. There is increasing recognition that, alongside published works meeting traditional standards, there are research papers of questionable quality, often with insufficient detail to understand the potential biases and limitations of the study's conclusions. Thus, a movement toward systematic, checklist-based disclosure of important methodological details began with the **CONSORT (Consolidated Standards of Reporting Trials) Statement**[33] of guidelines for reporting randomized trials and has blossomed into the **EQUATOR (Enhancing the QUAlity and Transparency Of health Research) network**,[34] which serves as a central point for identifying reporting guidelines for a wide range of study types.

SUMMARY OF REPORTING GUIDELINES

The rapid adoption of these guidelines by all major journal publishing houses validates well-described concerns that medical literature is of variable (and sometimes poor) quality,[35,36] and that standardized reporting improves the quality and transparency—and therefore, the reliability and value—of medical studies. A list of the relevant guidelines for several major study types is in Table 90.1.

Importantly, even if authors do not use a study quality tool to structure their manuscripts, readers can still use the guidelines to understand where important information may have been omitted, and judge for themselves whether the reported results are reliable (Box 90.1).

REGISTRATION OF CLINICAL TRIALS AND STUDY PROTOCOLS

Guidelines to assess the quality of completed, published research, while helpful, do not address the full spectrum of bias. There is a growing movement to register clinical trials (to prevent selective publication of results that confirm a particular, preferred hypothesis, or to prevent protocol changes that may obscure or invalidate the results of a study) and, increasingly, to register protocols for observational studies as well. Uptake has been better for prospective cohort studies than retrospective studies. In 2014, PLOS Medicine updated its requirements that observational studies must: adhere to the appropriate EQUATOR checklist and requirements for data sharing; specify hypotheses

TABLE 90.1 Study Types and Their Corresponding Reporting Guidelines

Study Type	Reporting Guidelines
Randomized trials	CONSORT
Observational studies	STROBE
Systematic reviews	PRISMA
Case reports	CARE
Qualitative research	SRQR
Diagnostic/prognostic studies	STARD
Prediction models	TRIPOD
Quality improvement studies	SQUIRE
Economic studies	CHEERS
Preclinical animal studies	ARRIVE
Study protocols	SPIRIT
Systematic review and metaanalysis protocols	PRISMA-P
Clinical practice guidelines	AGREE

From equator-network.org: Reporting guidelines for main study types.

BOX 90.1 Steps to Ensure the Publication of High-quality Research

Before Publication

Encourage uniformity in how research is reported (common reporting standards)
Enforce transparency in the way research is conducted (study registration, publication of study protocols or evolving methodological notes collected during data analysis, data sharing)
Pre-prints: Deposition in a repository to collect public comments (and criticism) prior to submission for peer review

In the Review Process

Incentivize reviewers to review critically and constructively (open peer review, publication of reviews along with a completed manuscript)
Incentivize journals to commit to publish research based on methodological rigor, regardless of the findings which are not known or not disclosed at the time of acceptance ("Registered Reports")

After Publication

Continue to accept public commentary in an open forum (e.g., PubPeer)
Publicly flag articles that are under investigation, and promptly retract those that are later found not to adhere to rigorous practices
Encourage replication studies linked to the original work
Reduce disincentives (i.e., normalize the practice) for authors to identify and retract their own articles where findings are due to methodological errors or otherwise cannot be replicated

and planned analyses, document the actual analyses, and explain any differences between the planned and actual analyses; and share the protocol for any prospective study.[37] Rigorous requirements like these are not yet widespread, to the detriment of the quality of the medical literature.

Interpreting the Analysis of a Study

For many, the "statistical analysis" section of a quantitative research study is quickly skimmed on the way to the more interesting results and discussion section. In some situations, like those with a fairly straightforward and well-standardized analytic plan, as with a simple randomized controlled trial, this may be an acceptable time-saving practice. However, the authors' experience is that often, readers do not have the training to interpret a moderately complex analytic plan, even with careful reading. In reality, gaining the training to do this with facility is not practical for most consumers of the medical literature, and the responsibility falls to journal reviewers to ensure that appropriate analytic plans are used in published work (see The Publication Process). We hope here to provide a brief primer and, importantly, equip readers with the knowledge to interpret the commonalities among research designs.

THE CONCERN OVER *P* < .05

Formerly codified as the altar upon which much quantitatively based medical research lives or dies, "$P < 0.05$" is increasingly viewed as an arbitrary threshold that says very little about clinical meaning, or even the likelihood that a finding is due to chance. Over mounting concern that *P*-value misuse was rampant, the American Statistical Society took the exceptionally unusual step of releasing a statement,[38] in March 2016, to help usher science into a "post $P < 0.05$ era."[39]

The statement addressed six key principles (Table 90.2), which should guide interpretation of *P*-values. We refer the interested reader to Greenland and colleagues' even more comprehensive list of ways to misinterpret a *P*-value; they list 25 and provide a helpful introduction to the philosophy behind *P*-values that is accessible to, and in fact targeted toward, non-statisticians.[40]

At a minimum, reporting of a statistical test should include both the *P*-value and a measure of **effect size,** such as an odds ratio or an absolute difference, and **uncertainty,** such as a confidence interval. A measure of absolute or relative difference yields the most likely magnitude and direction of the effect (based on model choices and assumptions). The confidence interval indicates the degree of precision in that estimate. Without even invoking a *P*-value, the reader now understands the effect size and uncertainty in the measurement and may decide, for himself or herself, whether the results aid understanding of a particular phenomenon. *P*-values may soon find themselves left out of the conversation.

REDUCING BIAS IN STUDY DESIGN

The analysis plan of a study depends, first, on what study design was chosen. Simplistically, certain study designs, such as a randomized, controlled trial, address potential bias introduced by baseline differences between the comparison and experimental groups by **randomization**, and others must control for that bias in other ways.

Discussed more thoroughly elsewhere in this book (see Chapter 89), randomization is a process used by prospective trials to attempt to control for bias between two or more comparison groups. In its ideal application, it results in a balance between measured and unmeasured factors between the two groups, such that the only difference is the receipt, or nonreceipt, of the treatment of interest. Randomization is the only design that is capable of definitively addressing **unmeasured confounders,** explaining why randomized controls are regarded as a high standard of evidence.

Randomization is not perfect, though. By its very nature, it is possible that certain covariates will (randomly) not be well balanced between the groups. Imbalance in measured covariates can also imply unmeasurable imbalance in those covariates which are unknown. This tends to be a bigger problem in small trials, where a small sample size is left unable to compensate for random variation. One potential

TABLE 90.2 American Statistical Society Statement on *P*-values

Six Principles	Interpretation
P-values can indicate how incompatible the data are with a specified statistical model	If the model is clearly incompatible, the *P*-value will often be small (e.g., because model assumptions are not met). A small *P*-value does not imply that the chosen model is relevant to the data.
P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone	The underlying analytic choices—which may be extremely complex—contribute as much to the relevance of the analysis as the model choice and the statistical test that is reported
Scientific conclusions and business or policy decisions should not be based only on whether a *P*-value passes a specific threshold	A conclusion is no more or less true at $P = .04$ than it is at $P = .06$; thresholds are arbitrary
Proper inference requires full reporting and transparency	A *P*-value is the product of its analytic plan and the data itself; therefore, the full output of all tested hypotheses (including those with *P*-values deemed "not significant") should be presented to prevent selective or biased reporting of statistically significant findings
A *P*-value, or statistical significance, does not measure the size of an effect or the importance of a result	Subtle and clinically irrelevant differences can have a tiny *P*-value, especially with highly precise measurements and/or a large sample size, and clinically important differences can be "nonsignificant" if the measurements are imprecise or the sample size is small. Statistical significance is not an indicator of clinical relevance
By itself, a *P*-value does not provide a good measure of evidence regarding a model or hypothesis	A *P*-value is inseparable from the details of its methodology and, without context, gives inadequate information about effect size and uncertainty so as to be essentially uninterpretable

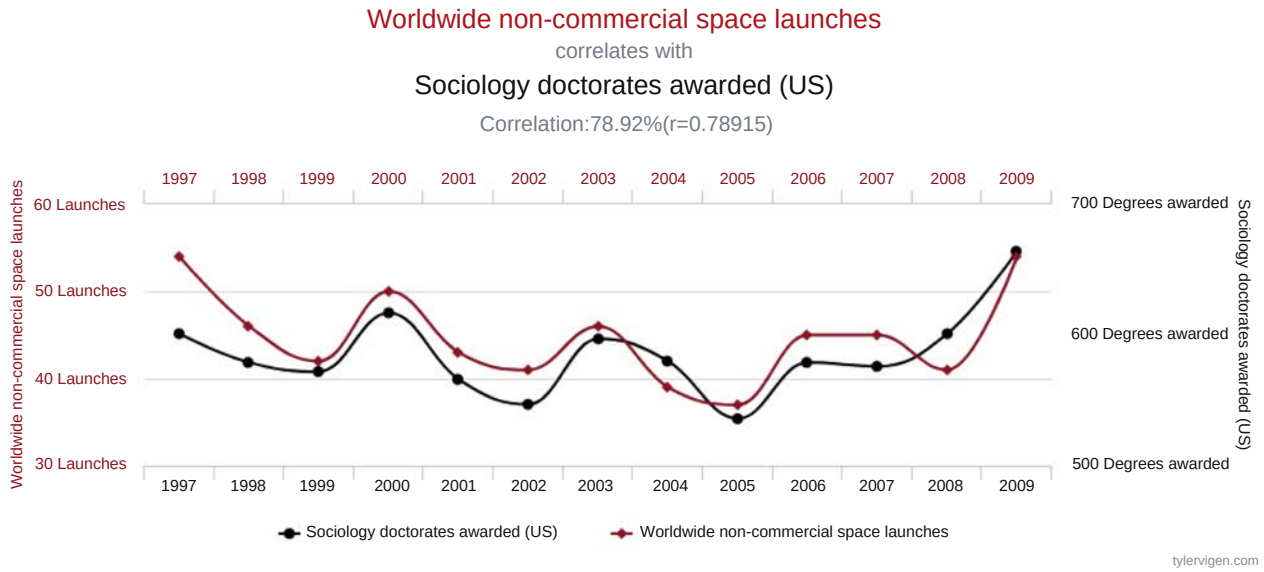

Fig. 90.4 A strong, but spurious, correlation. (From Vigen T. *Spurious Correlations*. New York: Hachette Books; 2015.)

option is to perform **stratified randomization**, where trial participants are first **stratified** by an important clinical predictor (e.g., smoking status), and then randomized to ensure that the groups are balanced on smoking status.

If a "treatment" or characteristic cannot be randomly assigned—for example, childhood socioeconomic status—a researcher could design a **matched** study, where each participant is matched to another who differs on receipt or nonreceipt of the treatment but is otherwise similar; such as, close in age, same gender, similar medical comorbidities. This method of balancing two groups is used in matched cohort studies and in propensity matching (whereby a probability of receiving a certain treatment is calculated, and participants with similar probability of treatment are matched according to whether they were or were not treated). The drawback is that even perfect balance of measured confounders does not ensure perfect balance of unmeasured confounders, which can be a serious source of bias.

A growing roster of increasingly complex statistical methods can be used to adjust for the individual influence of measured confounders mathematically, ultimately producing an estimate of an effect size **adjusted** for those measured confounders. Like any other statistical method, there are a litany of assumptions—some obvious, and some subtle—which must be satisfied for a model to produce accurate results. Like matching, however, adjustment cannot address unmeasured confounders, and there is an additional layer of complexity in that the modeling choices made by the researchers (e.g., which variables to include, and the strategy of building a model), which are rarely discussed in a manuscript, can have profound impact on the effect size estimate, the presence of "statistical significance," and even the direction of the purported effect. Common examples of adjustment methods are linear and logistic regression.

The fundamental differences between randomization and other ways of avoiding bias is the theoretical possibility to adjust for unmeasured confounders. For this reason, observational studies, which are typically limited to matching or mathematical adjustment, rarely produce strong evidence of **causation**. Rather, they can demonstrate an **association**, or a **correlation**—that is, that an outcome is more frequent after a certain treatment—but cannot prove that the outcome is caused by that treatment. This has a major impact on the language which can (or should) be used to describe the findings, and unfortunately, studies demonstrating correlation often find that the "translation" by popular media conflates *correlation* and *causation.*

CORRELATION VERSUS CAUSATION

More humorously, an entire website dedicated to the notion that **correlation does not imply causation** has identified mathematically implausible links between, for example, the number of noncommercial space launches and sociology doctorates awarded in the United States (Fig. 90.4).[41] The absurdness of this website serves to emphasize the importance of understanding when causal language is appropriate, and when relationships are only observed as associations. Furthermore, the underlying mechanism generating these correlations is that of multiple comparisons: the author uses publicly available data sources, along with a computer-based data mining algorithm, to test millions of correlations and identify the ones that are "statistically significant." The parallels between his methodology and that of unscrupulous, *P*-value-focused researchers are, unfortunately, clear.

DATA-DRIVEN ANALYTIC METHODS

The backbone of medical research is the hypothesis-driven study of medical outcomes. Understanding medical research methods and the underlying hypothesis allows us to determine which published work is hypothesis-driven, and which is not (often disparagingly referred to as "data dredging"). However, an emerging discipline of data-driven analytics, rooted in computer science rather than traditional epidemiology or medicine, is increasingly showing promise as a way to manage and glean insights from the enormous volumes of data that are collected through clinical work. Many of the analytic approaches look for patterns in data, without regard to known clinical

hypotheses or physiologically probable correlations. Traditional statistical concerns, like power, adjustment for multiple comparisons, nonrandom missing or erroneous data, and selective pursuit of "statistically promising" correlations, are difficult to frame in the context of this emerging methodology.

Artificial intelligence (AI), **machine learning**, and **deep learning** all refer to methods of data-driven analysis incorporating a learning algorithm: computer algorithms that change based on the data they receive.[42] The inherent complexity of these methods and the "learning" —by which an algorithm changes over time from how it was initially programmed to an evolution based on received data—make it extremely difficult to critically evaluate the quality of research that uses these analytic techniques. Usual parameters, such as *P*-values, confidence intervals, and point estimates may no longer apply. Metrics for understanding traditional "research quality" fail, but as these techniques increasingly gain acceptance, anesthesiology must continue to evolve its understanding of what data-driven methods may offer.

The Dark Side of the Medical Literature

In 2005, John Ioannidis published a seminal paper succinctly and provocatively titled, "Why Most Published Research Findings Are False,"[43] which helped spur the explosion of initiatives to improve the quality of research publications detailed earlier (see Tools to Ensure Study Quality). Two decades later, how have we done? A study attempting to replicate 100 high-profile research papers published in 2008 from three psychology journals depressingly concluded that only 39% could be replicated.[44] There is no reason to suspect disciplines other than psychology are not equally affected. Dr. Ioannidis's thesis holds that most published research findings are, apparently, false.[43]

INCENTIVES TO PUBLISH POOR-QUALITY RESEARCH

For Researchers

The incentives for researchers seem obvious. High academic productivity (i.e., publication) is explicitly or implicitly expected as a condition for ongoing funding, promotion, tenure, and compensation. Assuming that poor-quality research is not recognized as poor (for reasons of deliberate or unconscious exclusion of results that do not fit a hypothesis, undisclosed or unrecognized bias, successfully disguised research fraud, and so on), the number of publications, and the quality of the journals in which they are published, provides a quick way to decide whether a researcher is an expert or a dilettante.

But beyond academic prestige, some countries explicitly award cash bonuses for high-profile publications in journals such as *Science*[45] and *Nature*.[46] Availability of a government-initiated cash reward for successful publication is associated with a 46% increase in submissions to *Science*, although the authors of the study infer that, because of lower acceptance rates, cash incentives also encourage the submission of lower-quality research.[47] A recent survey of the monetary reward system in China revealed the magnitude of these potential rewards: cash incentives for publication in the most prestigious journals can be up to 20 times the annual salary of a university professor.[48]

Another incentive has to do with the conclusions of the research itself, and the likelihood of publication. Strong and surprising associations tend to be easier to publish in top journals, while subtle, or unsurprising, or complex/interdependent findings are more likely to be relegated to a small journal, resulting in less reward for the investigator. A "surprising" association implies that the pre-test probability, the likelihood of finding that association, given what we knew before the study, was low. But as Ioannidis shows, that "surprising" association only increases the post-test probability of an improbable association slightly, because of the high likelihood that the newly-demonstrated association is due to chance and/or bias.[43] It remains probable that the finding is wrong, particularly if flexible data analysis methods have been used and the authors have, intentionally or not, engaged in selective reporting. Nonetheless, a novel, statistically significant effect is likely to find a home in a journal, and mathematical modeling studies have shown how spurious findings can become "canonized" into facts using plausible values for publication probabilities that depend upon whether the finding is positive or negative, with bias against publication of negative findings.[49]

Furthermore, conducting adequately powered experiments with low risk of bias requires effort. Lower effort, as in small trials, with high risk of bias, is paradoxically more likely to generate novel positive results, and novel positive results are more likely to result in the reward of publication. Even without conscious bias a low-effort approach can generate the trappings of academic success while adding little to reproducible, meaningful medical research.

For Medical Journals

In 1980, a 101-word letter to the editor titled "Addiction Rare in Patients Treated with Narcotics" was published in the high-impact *NEJM*. It disclosed no methodological details, provided no information about the included patients, and used the "non-standard" interpretation of "reasonably well documented addiction," which was reported in 4 of roughly 12,000 patients receiving a narcotic prescription. The letter went on to be cited more than 600 times and may have provided foundational support to a narrative promoted by opioid manufacturers that was used to mislead prescribers, users, and regulators of opioid medications.[50] The original letter is now modified by an editor's note on the *NEJM* website, that "For reasons of public health, readers should be aware that this letter has been 'heavily and uncritically cited' as evidence that addiction is rare with opioid therapy."[51]

Bad science, particularly when it finds itself supporting a position that brings frequent references from follow-on work, can nonetheless reward its publishing journal with an improvement in ***impact factor***, a heavily criticized but still universal method of ranking the "best" journals. The impact factor is the average number of citations per paper over a 2-year interval, and thus, frequently cited papers repay a journal with an increase in impact factor.

Conversely, replication studies are rarely cited, and thus journals are not incentivized to publish them. Since it is an average, a journal's impact factor clearly does not directly reflect the quality of an individual paper, nor can it be used to distinguish high-quality research from highly sensational work in an explosive field, or citations resulting from well-founded post-publication criticism from the research community (as in the case of the *NEJM* letter, noted previously, which we have deliberately avoided citing in this chapter).

MISCONDUCT IN THE PUBLICATION PROCESS

The availability of the Internet for the communication of research findings has revolutionized the academic publishing business. However, the fundamental components of rigorous peer review, editorial decisions, and publishing services remain largely unchanged. It is entirely possible, and perhaps desirable, to publish well-conducted research in an online-only format, after proper peer review, and it is also completely appropriate for open access publications to charge fees that support the costs of the review and publication process and the journal's overhead (in the absence of subscription fees). However, there has been exploitation of the open-access movement: **Predatory journals**, often styled as "open access," take advantage of the publication fees required by legitimate open access journals, sometimes on the order of several thousand dollars, to "publish" an article without appropriate peer review or editorial oversight. Journals may promise an exceptionally rapid peer review process (easily accomplished when no peer review takes place), offer low barriers to publication such as partial publication cost waivers, mention publication costs late in the review process after the manuscript has been accepted, or invite editorials or review articles with no specified topic.

However, it is very difficult to distinguish a journal's quality based on those characteristics alone, as legitimate journals may conduct rapid peer review (with reviewers motivated to act with a short turn-around time), offer page publication waivers for specified situations, and invite articles on a wide variety of topics. An attempt has been made to generate and maintain a list of predatory journals and criteria to identify them.[52] Unfortunately, the distinction is not black-and-white, the list of criteria may be helpful, as may a website available at thinkchecksubmit.org,[53] founded by a coalition of scholarly organizations (including *BioMed Central*[54] and the *Committee on Publication Ethics*[55]), which provides authors with a checklist to help identify trustworthy journals. Fundamentally, consumers of the medical literature must be aware that a published article no longer implies any degree of peer review in many of these predatory journals.

There are also obvious potential sources for misconduct in the review process, such that the *Committee on Publication Ethics* maintains detailed flow sheets to assist journal editors in addressing allegations of misconduct, such as plagiarism, duplicate publication, data fabrication, "ghost" or "gift" authorship, and reviewer appropriation of an author's ideas.[56] These have been well described, are clearly unethical, and continue to evolve: a recent "new" source of misconduct is the spoofing of reviewers, where an author recommends reviewers either under names of real scientists or pseudonyms, but lists spurious email addresses that go to colleagues or to the author him/herself. The author then completes the "peer review," recommending minimal or no changes. This type of fraud can be extremely difficult to detect, as are other ways of "gaming the system," which have no doubt evolved even further since *Nature* published an initial exposé in 2014.[57]

In 2017, a Discover Magazine blog called Neuroskeptic published a post about a Star Wars-themed spoof manuscript's success in the dark world of predatory journals.[58] The nonsensical paper—"an absurd mess of factual errors, plagiarism and movie quotes," in the words of its author—about midi-chlorians (a fictional intracellular entity closely linked to the transmission of Jedi powers) was accepted by the *American Journal of Medical and Biological Research* (although a fee was subsequently requested prior to publication, which the blogger did not pay) and was published by the *International Journal of Molecular Biology: Open Access*, *Austin Journal of Pharmacology and Therapeutics*, and the *American Research Journal of Biosciences*. The publications were withdrawn after the spoof was revealed.

Even a cursory reading of the paper would have sufficed for someone unfamiliar with Star Wars to identify the paper as unfocused and unclear, with bizarre colloquialisms; of course, it is also written about a fictional organelle, which should have been immediately recognized by any reviewer with appropriate experience in cell biology. Nonetheless, beyond the manuscript's successful publication in three journals, the fictional first author of the "manuscript," Dr. Lucas McGeorge, was shortly thereafter offered an unsolicited invitation to join the editorial board of another research journal.[58] This anecdote proves, to a distressing degree, the existence of these unscrupulous journals, and how distortion of the medical literature could easily occur.

EVOLVING MECHANISMS TO DETECT RESEARCH MISCONDUCT

How might we align incentives in the manuscript preparation, review, and post-publication process to achieve checks on the amount of low-quality research that is published?[59] Currently, the supposedly "self-correcting" nature of the medical literature is in question.[60] Addressing this may require fundamental changes to the way the medical community views research—its generation and its consumption—but refusal to change will undermine public trust in a massive, productive, and necessary enterprise.

In 2012, a nearly two-decades-long career in anesthesiology research built on fabrication was exposed,[61] ten years after the author's work was first publicly called into question.[62] One hundred and sixty-eight randomized trials performed by Yoshitaka Fujii and colleagues were statistically compared against similar trials from other authors. Although the underlying statistical analysis may not have been optimal, Carlisle demonstrated that Fujii's reported distributions of participant characteristics were extremely unlikely to be observed in nature—values were too reliably close to population means, failing to display expected variability.[63] An investigation by Fujii's home institution uncovered widespread data fabrication, and ultimately

concluded that "it is as if someone sat at a desk and wrote a novel about a research idea."[64]

As a result, 172 of Fujii's papers have been retracted. The stunning scope of the fraud is eclipsed by the even more stunning duration of his misconduct: at least 10 years of fabrication followed by another decade of public speculation without actionable results. High-profile misconduct cases like this have prompted calls for widespread adoption of automated methods to identify suspicious work both pre- and post-publication.[65] However, those calls have been met with an equally loud resistance, on statistical and ethical grounds. The criticisms focus on the method's inadequate consideration of expected interdependence among study participant baseline variables among other methodological problems, and highlight the potential stigmatization of honest researchers that could stem from an inadequately nuanced approach to questions about academic integrity.[66,67] Ultimately, as in the case of plagiarism-checking software (in widespread use in medical publishing), careful and dispassionate human interpretation will prove necessary as these techniques are refined and gain greater acceptance.

Conclusions

Medical knowledge is evolving at a more rapid pace than ever before. Understanding how to approach a published study, from its design to the quality of its conclusions, is necessary for individual and systems-level care improvement. Alertness to known predictors of incomplete or inaccurate information will help clinicians and scientists continue to generate useful knowledge and provide the best, evidence-based anesthesia care.

Complete references available online at expertconsult.com.

References

1. Sackett DL. *Can Med Assoc J*. 1982;126(12):1373.
2. Burls A. What is critical appraisal? *What is...? Series*. 2009:1–8.
3. Smith R, Rennie D. *JAMA*. 2014;311(4):365–367. https://doi.org/10.1001/jama.2013.286182.
4. Oxman AD, et al. *JAMA*. 1993;270(17):2093–2095.
5. Gordon G, et al. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. 3rd ed. McGraw-Hill Education; 2015.
6. The new england journal of medicine. https://www.nejm.org/.
7. Journal of the american medical association. https://jamanetwork.com/journals/jama/currentissue.
8. The BMJ. https://www.bmj.com/. Accessed May 8, 2018.
9. The lancet. https://www.thelancet.com/. Accessed May 8, 2018.
10. Anesthesiology. http://anesthesiology.pubs.asahq.org/journal.aspx. Accessed May 8, 2018.
11. British Journal of Anaesthesia. https://bjanaesthesia.org/. Accessed May 8, 2018.
12. Anaesthesia & Analgesia. https://journals.lww.com/anesthesia-analgesia/pages/default.aspx. Accessed May 8, 2018.
13. University of Illinois at Chicago's Library of the Health Sciences at Peoria. Evidence based medicine: levels of evidence. https://researchguides.uic.edu/c.php?g=252338&p=3950157. Accessed May 7, 2018.
14. PubMed. https://www.ncbi.nlm.nih.gov/pubmed/. Accessed May 8, 2018.
15. National center for biotechnology information. https://www.ncbi.nlm.nih.gov/. Accessed May 8, 2018.
16. U.S. National library of medicine. https://www.nlm.nih.gov/. Accessed May 8, 2018.
17. National institutes of health. https://www.nih.gov/. Accessed May 8, 2018.
18. MEDLINE: description of the database. https://www.nlm.nih.gov/bsd/medline.html. Accessed May 8, 2018.
19. Cochrane database of systematic reviews. http://www.cochranelibrary.com/cochrane-database-of-systematic-reviews/index.html. Accessed May 8, 2018.
20. UptoDate. https://www.uptodate.com/home. Accessed May 8, 2018.
21. WebMD. https://www.webmd.com/. Accessed May 8, 2018.
22. Twitter. https://www.twitter.com. Accessed May 8, 2018.
23. Facebook. https://www.facebook.com/. Accessed May 8, 2018.
24. Wikipedia. https://www.wikipedia.org/. Accessed May 8, 2018.
25. Evidence-based medicine: resources by levels of evidence. https://libguides.cmich.edu/cmed/ebm/pyramid. Accessed May 8, 2018.
26. Murad MH, et al. *Evid Based Med*. 2016;21(4):125–127.
27. Salvador-Carulla L, et al. *Epidemiol Psychiatr Sci*. 2017;26(2):105–114.
28. Walach H, et al. *BMC Med Res Methodol*. 2006;6:29.
29. Berlin JA, Golub RM. *JAMA*. 2014;312(6):603–605.
30. Roland M, Torgerson DJ. *BMJ*. 1998;316(7127):285.
31. Semler MW, et al. *N Engl J Med*. 2018;378(9):829–839.
32. Self WH, et al. *N Engl J Med*. 2018;378(9):819–828.
33. Consort statement. http://www.consort-statement.org/. Accessed May 8, 2018.
34. Enhancing the QUAlity and transparency of health research. http://www.equator-network.org/. Accessed May 8, 2018.
35. Thakur A, et al. *J Pediatr Surg*. 2001;36(8):1160–1164.
36. Chan AW, Altman DG. *Lancet*. 2005;365(9465):1159–1162.
37. PLOS Medicine Editors. *PLoS Med*. 2014;11(8):e1001711.
38. Wasserstein RL, Lazar NA. *The American Statistician*. 2016;70(2):129–133.
39. American Statistical Organization. www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf. Accessed May 8, 2018.
40. Greenland S, et al. *Eur J Epidemiol*. 2016;31(4):337–350.
41. Vigen T. Spurious correlations. http://www.tylervigen.com/spurious-correlations. Accessed May 8, 2018.
42. LeCun Y, et al. *Nature*. 2015;521(7553):436–444.
43. Ioannidis JP. *PLoS Med*. 2005;2(8):e124.
44. Open Science Collaboration. *Science*. 2015;349(6251):aac4716. https://doi.org/10.1126/science.aac4716.
45. Science. http://www.sciencemag.org/. Accessed May 8, 2018.
46. Nature. https://www.nature.com/nature/. Accessed May 8, 2018.
47. Franzoni C, et al. *Science*. 2011;333(6043):702–703.
48. Quan W, et al. *Aslib Journal of Info Mgmt*. 2017;69(5):486–502.
49. Nissen SB, et al. *Elife*. 2016;5.
50. Leung PTM, et al. *N Engl J Med*. 2017;376(22):2194–2195.
51. Addiction rare in patients treated with narcotics, with editor's note. https://www.nejm.org/doi/10.1056/NEJM198001103020221. Accessed May 8, 2018.
52. Beall J. Criteria for determining predatory open-access publishers. https://scholarlyoa.files.wordpress.com/2012/11/criteria-2012-2.pdf. Accessed August 5, 2019.
53. Think Check. Submit. https://thinkchecksubmit.org/. Accessed May 8, 2018.
54. BioMed central. https://www.biomedcentral.com/. Accessed May 8, 2018.
55. Committee on Publication Ethics (COPE). https://publicationethics.org/. Accessed May 8, 2018.
56. Committee on publication ethics flowcharts. https://publicationethics.org/resources/flowcharts. Accessed May 8, 2018.
57. Ferguson C, et al, Oransky I. *Nature*. 2014;515(7528):480–482.
58. Neuroskeptic. *Predatory Journals hit by 'Star Wars' Sting*. 2017.
59. Gorman DM, et al. *Sci Eng Ethics*. 2017.
60. Allison DB, et al. *Nature*. 2016;530(7588):27–29.
61. Carlisle JB. *Anaesthesia*. 2012;67(5):521–537.
62. Kranke P, et al. *Anesth Analg*. 2000;90(4):1004–1007.
63. Carlisle JB, et al. *Anaesthesia*. 2015;70(7):848–858.
64. Anesthesiologist fabricates 172 papers. https://www.the-scientist.com/?articles.view/articleNo/32312/title/Anesthesiologist-Fabricates-172-Papers/. Accessed May 7, 2018.
65. Loadsman JA, McCulloch TJ. *Anaesthesia*. 2017;72(8):931–935.
66. Kharasch ED, Houle TT. *Anesthesiology*. 2017;127(5):733–737.
67. Mascha EJ, et al. *Anesth Analg*. 2017;125(4):1381–1385.

References

1. Sackett DL. How to read clinical journals. *Can Med Assoc J*. 1982;126(12):1373.
2. Burls A. *What is Critical Appraisal?* 2009:1–8.
3. Smith R, Rennie D. Evidence-based medicine—an oral history. *JAMA*. 2014;311(4):365–367. https://doi.org/10.1001/jama.2013.286182.
4. Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. The evidence-based medicine working group. *JAMA*. 1993;270(17):2093–2095.
5. Guyatt Gordon, Rennie Drummond, Meade Maureen O, Cook Deborah J. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. 3rd ed. McGraw-Hill Education; 2015.
6. The New England Journal of Medicine. https://www.nejm.org/.
7. Journal of the American Medical Association. https://jamanetwork.com/journals/jama/currentissue.
8. The BMJ. https://www.bmj.com/. Accessed May 8, 2018.
9. The Lancet. https://www.thelancet.com/. Accessed May 8, 2018.
10. Anesthesiology. http://anesthesiology.pubs.asahq.org/journal.aspx. Accessed May 8, 2018.
11. British Journal of Anaesthesia. https://bjanaesthesia.org/. Accessed May 8, 2018.
12. Anaesthesia & Analgesia. https://journals.lww.com/anesthesia-analgesia/pages/default.aspx. Accessed May 8, 2018.
13. University of Illinois at Chicago's Library of the Health Sciences at Peoria. Evidence based medicine: levels of evidence. https://researchguides.uic.edu/c.php?g=252338&p=3950157. Accessed May 7, 2018.
14. PubMed. https://www.ncbi.nlm.nih.gov/pubmed/. Accessed May 8, 2018.
15. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/. Accessed May 8, 2018.
16. U.S. National Library of Medicine. https://www.nlm.nih.gov/. Accessed May 8, 2018.
17. National Institutes of Health. https://www.nih.gov/. Accessed May 8, 2018.
18. MEDLINE: description of the database. https://www.nlm.nih.gov/bsd/medline.html. Accessed May 8, 2018.
19. Cochrane Database of Systematic Reviews. http://www.cochranelibrary.com/cochrane-database-of-systematic-reviews/index.html. Accessed May 8, 2018.
20. UptoDate. https://www.uptodate.com/home. Accessed May 8, 2018.
21. WebMD. https://www.webmd.com/. Accessed May 8, 2018.
22. Twitter. https://www.twitter.com. Accessed May 8, 2018.
23. Facebook. https://www.facebook.com/. Accessed May 8, 2018.
24. Wikipedia. https://www.wikipedia.org/. Accessed May 8, 2018.
25. Evidence-based medicine: resources by levels of evidence. https://libguides.cmich.edu/cmed/ebm/pyramid. Accessed May 8, 2018.
26. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. *Evid Based Med*. 2016;21(4):125–127. https://doi.org/10.1136/ebmed-2016-110401.
27. Salvador-Carulla L, Lukersmith S, Sullivan W. From the EBM pyramid to the greek temple: a new conceptual approach to guidelines as implementation tools in mental health. *Epidemiol Psychiatr Sci*. 2017;26(2):105–114. https://doi.org/10.1017/S2045796016000767.
28. Walach H, Falkenberg T, Fonnebo V, Lewith G, Jonas WB. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol*. 2006;6:29.
29. Berlin JA, Golub RM. Meta-analysis as evidence: building a better pyramid. *JAMA*. 2014;312(6):603–605. https://doi.org/10.1001/jama.2014.8167.
30. Roland M, Torgerson DJ. What are pragmatic trials? *BMJ*. 1998;316(7127):285.
31. Semler MW, Self WH, Wanderer JP, et al. Balanced crystalloids versus saline in critically ill adults. *N Engl J Med*. 2018;378(9):829–839. https://doi.org/10.1056/NEJMoa1711584.
32. Self WH, Semler MW, Wanderer JP, et al. Balanced crystalloids versus saline in noncritically ill adults. *N Engl J Med*. 2018;378(9):819–828. https://doi.org/10.1056/NEJMoa1711586.
33. Consort statement. http://www.consort-statement.org/. Accessed May 8, 2018.
34. Enhancing the quality and transparency of health research. http://www.equator-network.org/. Accessed May 8, 2018.
35. Thakur A, Wang EC, Chiu TT, et al. Methodology standards associated with quality reporting in clinical studies in pediatric surgery journals. *J Pediatr Surg*. 2001;36(8):1160–1164.
36. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet*. 2005;365(9465):1159–1162.
37. PLOS Medicine Editors. Observational studies: getting clear about transparency. *PLoS Med*. 2014;11(8):e1001711. https://doi.org/10.1371/journal.pmed.1001711.
38. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. *The American Statistician*. 2016;70(2):129–133. https://doi.org/10.1080/00031305.2016.1154108.
39. American Statistical Organization. American Statistical Association releases statement onstatistical significance and p-values. www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf. Accessed May 8, 2018.
40. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol*. 2016;31(4):337–350. https://doi.org/10.1007/s10654-016-0149-3.
41. Vigen T. Spurious correlations. http://www.tylervigen.com/spurious-correlations. Accessed May 8, 2018.
42. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444. https://doi.org/10.1038/nature14539.
43. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.
44. Open Science Collaboration. PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science*. 2015;349(6251):aac4716. https://doi.org/10.1126/science.aac4716.
45. Science. http://www.sciencemag.org/. Accessed May 8, 2018.
46. Nature. https://www.nature.com/nature/. Accessed May 8, 2018.
47. Franzoni C, Scellato G, Stephan P. Science policy. Changing incentives to publish. *Science*. 2011;333(6043):702–703. https://doi.org/10.1126/science.1197286.
48. Quan W, Chen B, Shu F. Publish or impoverish: an investigation of the monetary reward system of science in China (1999-2016). *Aslib Journal of Info Mgmt*. 2017;69(5):486–502.
49. Nissen SB, Magidson T, Gross K, Bergstrom CT. Publication bias and the canonization of false facts. *Elife*. 2016;5. https://doi.org/10.7554/eLife.21451.
50. Leung PTM, Macdonald EM, Stanbrook MB, Dhalla IA, Juurlink DN. A 1980 letter on the risk of opioid addiction. *N Engl J Med*. 2017;376(22):2194–2195. https://doi.org/10.1056/NEJMc1700150.
51. Addiction rare in patients treated with narcotics, with editor's note. https://www.nejm.org/doi/10.1056/NEJM198001103020221. Accessed May 8, 2018.
52. Beall J. Criteria for determining predatory open-access publishers. https://scholarlyoa.files.wordpress.com/2012/11/criteria-2012-2.pdf. Accessed May 8, 2018.
53. Think. Check. Submit. https://thinkchecksubmit.org/. Accessed May 8, 2018.
54. BioMed central. https://www.biomedcentral.com/. Accessed May 8, 2018.
55. Committee on Publication Ethics (COPE). https://publicationethics.org/. Accessed May 8, 2018.
56. Committee on Publication Ethics flowcharts. https://publicationethics.org/resources/flowcharts. Accessed May 8, 2018.
57. Ferguson C, Marcus A, Oransky I. Publishing: the peer-review scam. *Nature*. 2014;515(7528):480–482. https://doi.org/10.1038/515480a.
58. Neuroskeptic. *Predatory Journals Hit By 'Star Wars' Sting*. 2017.
59. Gorman DM, Elkins AD, Lawley M. A systems approach to understanding and improving research integrity. *Sci Eng Ethics*. 2017. https://doi.org/10.1007/s11948-017-9986-z.
60. Allison DB, Brown AW, George BJ, Kaiser KA. Reproducibility: a tragedy of errors. *Nature*. 2016;530(7588):27–29. https://doi.org/10.1038/530027a.
61. Carlisle JB. The analysis of 168 randomised controlled trials to test data integrity. *Anaesthesia*. 2012;67(5):521–537. https://doi.org/10.1111/j.1365-2044.2012.07128.x.

62. Kranke P, Apfel CC, Roewer N, Fujii Y. Reported data on granisetron and postoperative nausea and vomiting by Fujii et al. are incredibly nice! *Anesth Analg*. 2000;90(4):1004–1007.
63. Carlisle JB, Dexter F, Pandit JJ, Shafer SL, Yentis SM. Calculating the probability of random sampling for continuous variables in submitted or published randomised controlled trials. *Anaesthesia*. 2015;70(7):848–858. https://doi.org/10.1111/anae.13126.
64. Anesthesiologist fabricates 172 papers. https://www.the-scientist.com/?articles.view/articleNo/32312/title/Anesthesiologist-Fabricates-172-Papers/. Accessed May 7, 2018.
65. Loadsman JA, McCulloch TJ. Widening the search for suspect data—is the flood of retractions about to become a tsunami? *Anaesthesia*. 2017;72(8):931–935. https://doi.org/10.1111/anae.13962.
66. Kharasch ED, Houle TT. Errors and integrity in seeking and reporting apparent research misconduct. *Anesthesiology*. 2017;127(5):733–737. https://doi.org/10.1097/ALN.0000000000001875.
67. Mascha EJ, Vetter TR, Pittet JF. An appraisal of the Carlisle-Stouffer-Fisher method for assessing study data integrity and fraud. *Anesth Analg*. 2017;125(4):1381–1385. https://doi.org/10.1213/ANE.0000000000002415.