Demystifying the Medical Literature

Christopher Greeley

ABSTRACT

The published literature is not just the ongoing record of current medical and scientific knowledge; it is a record of the past and can give an eye toward future knowledge. Reading the published literature can give a view of the evolution of knowledge on a particular question, the growth of a discipline, the identification of new diseases, and the refinement of diagnostic tests. The reality is that most busy physicians read only the abstract of an article. The purpose of this article is to place published medical literature into a context and to provide some considerations for critically evaluating articles. This paper will provide historic background of evidence-based medicine and medical publications. Specific strategies for critical literature appraisal are highlights and pitfalls to avoid are outlined. *Acad Forensic Pathol.* 2016 6(4): 556-567

AUTHOR

Christopher Greeley MD, Baylor College of Medicine - Pediatrics

Roles: Project conception and/or design, data acquisition, analysis and/or interpretation, manuscript creation and/or revision, approved final version for publication, accountable for all aspects of the work, principal investigator of the current study.

CORRESPONDENCE

Christopher Greeley MD, One Baylor Plaza, Houston TX 77030-3411, Greeley@bcm.edu

ETHICAL APPROVAL

As per Journal Policies, ethical approval was not required for this manuscript

STATEMENT OF HUMAN AND ANIMAL RIGHTS

This article does not contain any studies conducted with animals or on living human subjects

STATEMENT OF INFORMED CONSENT

No identifiable personal data were presented in this manuscript

DISCLOSURES & DECLARATION OF CONFLICTS OF INTEREST

The author, reviewers, editors, and publication staff do not report any relevant conflicts of interest

FINANCIAL DISCLOSURE

The author has indicated that he does not have financial relationships to disclose that are relevant to this manuscript

KEYWORDS

Forensic pathology, Evidence-Based Medicine, Literature appraisal

INFORMATION

ACADEMIC FORENSIC PATHOLOGY: THE OFFICIAL PUBLICATION OF THE NATIONAL ASSOCIATION OF MEDICAL EXAMINERS

©2017 Academic Forensic Pathology International • (ISSN: 1925-3621) • <https://doi.org/10.23907/2016.055>

Submitted for consideration on 5 Oct 2016. Accepted for publication on 21 Nov 2016

INTRODUCTION

The published literature is not just the ongoing record of current medical and scientific knowledge, it is a record of the past and can give an eye towards future knowledge. Reading the published literature can give a view of the evolution of knowledge on a particular question, the growth of a discipline, the identification of new diseases, and the refinement of diagnostic tests. The reality is that most busy physicians read only the abstract of an article (if the title is interesting enough to entice them to read further) and do not routinely read the vast majority of the paper. If the paper is read, most professionals tend to read an article from beginning to end as they would a novel, with only a rudimentary sense of some of the scaffolding on which a quality paper should be framed. The purpose of this article is to place published medical literature into a context and to provide some considerations for critically evaluating articles.

DISCUSSION

Evidence-Based Medicine

Modern critical literature appraisal was framed within the rise of the evidence-based medicine (EBM) movement. Evidence-based medicine is widely credited to have originated in McMaster University (Ontario, Canada) in the 1990s. Drs. David Sackett and Gordon Guyatt led the Evidence-Based Working Group (1), framing a new paradigm of patient care and medical scholarship. Per Dr. Sackett, EBM is “*the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients*” (2). Evidenced-based medicine blends the clinical experience and judgment of the practitioner with a systematic analysis and critical appraisal of the appropriate published literature. The Evidence-Based Working Group promoted a medical practice which transitioned from one based solely upon intuition, anecdote or supposition to one which incorporated the assessment of evidence collected through a systematic literature search and reflected that upon the clinical practice and experience of the physician. This paradigm shift required the practitioner to develop new

skills: systematic literature appraisal, evidence quality rating, and an evaluation of the quality of the study performed.

As there are a number excellent comprehensive resources available for a deeper description of EBM (3, 4), I will limit this discussion to only some of the underlying principles. The framework of EBM has five fundamental steps as outlined by Dr. Sackett (1, 5), which constitute the Critical Appraisal Exercise (1):

1) convert these information needs into answerable questions; 2) track down, with maximum efficiency, the best evidence with which to answer them (and making increasing use of secondary sources of the best evidence); 3) critically appraise that evidence for its validity (closeness to the truth) and usefulness (clinical applicability); 4) integrate the appraisal with clinical expertise and apply the results in clinical practice; and 5) evaluate one's own performance (5).

These five principles outline a process of understanding a clinical question, identifying the appropriate literature that can answer the question and appraising the evidence within the published studies to inform clinical care. It is apparent that EBM is largely dependent upon a critical appraisal of the literature to uncover the answer to the proposed clinical question.

There have been some notable criticisms of EBM and its framework. One of the earliest concerns, which arose at the beginning of the formulation of EBM, was that proponents of EBM were overconfident in their strength of its claims (6, 7). The early cautions were directed towards a discounting of unpublished works (negative study publication bias), as well as the overreliance on randomized controlled trials (RCT), the under appreciation of clinical experience, the unrealistic time required by a practitioner to perform a rigorous literature review, and the loss of the “art of medicine” (8). More recent criticisms of EBM centers around its algorithmic nature (9, 10). Thomas, quite convincingly, argues that the unwavering commitment to RCT as a pinnacle of evidence suppresses large swaths of meaningful outcomes and research

contexts that are very meaningful to the complex nature of individual patient diagnosis and care (9). Additionally, van Baalen and Boon argue that the “messy” nature of clinical care is not suited for the “rule-based reasoning” promoted by EBM (10).

Journals

The first step in a critical appraisal of the medical literature is to appreciate that the journals themselves are not simply dispassionate blank slates. Journals and their editors often have a particular emphasis or perspective that is being promoted. It is important for the reader to be aware that journals, like anything that we read, can contain both high-quality and low-quality papers. While a journal’s impact factor (IF) is often cited as a measure of its quality, it has been known for decades that the IF has some significant flaws (11, 12). A journal’s IF is, basically, the rate at which research papers published are cited within the subsequent two years (12). Thus a journal with an IF of 5.0 means that the “average citable” paper within that journal is cited an average of five times in the two years after it is published. As is apparent, this number can, in no way, be a reflection of the quality of the journal or the papers it published, and does not reflect anything about a specific paper in front of the reader (13). For example, the importance of an influential or transformative paper (i.e., Watson and Crick’s 1953 description of the double helix structure of DNA [14]) may not have been recognized within the two years subsequent to its publication, and thus would not “count” towards the IF. Likewise, a provocative paper that contains fraudulent data (i.e., Wakefield et al.’s retracted 1998 paper proposing a link between the MMR [measles, mumps, rubella] vaccine and autism [15]) may be cited frequently, decrying its poor quality and fraudulent nature, but that would “count” towards the journal’s IF (Wakefield et al.’s paper (15) was cited 126 times in the subsequent two years after publication). It is also notable that most journals seek repeat submissions by productive authors. As most authors have a particular research trajectory or content area, and authors tend to cite their own prior work, journals would then benefit from self-citation.

One of the biggest issues confronting current scholars is the profound increase in the sheer volume of what is being published. This was most clearly demonstrated by Fraser and Dunstan (16). They reported that within cardiology, a new entrant into the field (in 2010) who wanted to read “everything” about echocardiography would have to read 156 661 articles. If they read five papers per hour for eight hours per day, for five days a week, for 50 weeks per year, it would take them over 11 years to get through “everything.” During this 11 years, an additional 82 142 papers would be published, requiring an additional eight years to get through. Not only is the volume of materials accelerating, there has been an increase in the rates of retractions reported by PubMed (17).

One reason for the increase in volume is the rise of open access (OA) journals. Open access journals provide free full-text access to the general public, with the costs associated for processing and publishing paid for by the author (18, 19). The genesis of OA publishing was coincident with the rise of the Internet and is based in the laudable goal to disseminate scientific knowledge quite broadly (18) and while there are a number of very high-quality OA journals (e.g., PLOS One and BioMed Central), it is important for the reader to be ever mindful of the growth of “predatory publishing.” Predatory publishing is when submissions are sought for publication with the goal being to solicit “publication” fees from the author (18, 20). Many in academia have received unsolicited emails requesting submissions to any number of journals and promising a rapid review and publication within one to two weeks. An investigation by the writer John Bohannon for *Science* reported on the network of shady publishers scattered across the globe (21). Bohannon created a journalistic sting by creating 304 spoof manuscripts which appeared to come out of an African University claiming to cure various forms of cancer. Despite being fictitious, not being scientifically valid (as determined by two external reviewers prior to the sting), and containing sham data, 157 journals accepted the manuscript for publication. Only 36 journals generated any peer review comments with 16 of those publishing the manuscript over the reviewer’s recommendations. In a recent commentary, Byard de-

scribes his experiences with predatory publishing and the potential negative impact it could have on forensic pathology (20). A recent estimate is that the number of OA journals equals that of conventional, subscription-based, journals (approximately 14 000 of each) (18). A “Black List” of predatory and suspect journals is maintained online at <https://scholarlyoa.com/publishers/>. In being mindful of the growth of OA publications, readers could search in more established sources of medical publications. The most well known is PubMed (22) (or PubMed Central, the repository for full-text articles). This is the online portal for the U.S. National Library of Medicine (NLM) within the National Institutes of Health (NIH). For a journal to be “indexed” (included into PubMed) it undergoes a quality review by NLM staff.

Peer Review

Given the changing landscape of the published medical literature, the historic reliance upon rigorous, high-quality peer review can no longer be relied upon. Peer review is a mechanism by which there is external review of a submitted manuscript by a qualified reviewer. As the volume of publications continues to explode, peer review has been an early victim. As a rule, all peer review is voluntary and uncompensated. While this is done in the spirit that dispassionate professionals will read a manuscript and, in some way, sanctify its contents as passing a threshold of quality. In reality, peer review being voluntary and uncompensated places the task at the bottom of the stack of things to do by busy professionals. Two complicating issues for the reader are that there exists no standard format, structure or content by which an external reviewer should perform their review, and who is chosen to review the manuscript is at the discretion of the editorial staff of the journal. Each journal has separate review processes, some of which are robust, and some of which are vague and rudimentary. As reviewers are not given any training or guidance, the quality of the review can suffer. This results in a great risk of high quality manuscripts being rejected (Type 2 error), and low quality manuscripts being accepted (Type 1 error) (18). This is counter to the foundation of the scientific enterprise.

An example of the potential flaws within peer review was most clearly demonstrated by Godlee et al. in 1998 (23). The authors took a paper that had already been accepted to the *British Medical Journal* (a high-quality, peer reviewed journal) and introduced eight weaknesses into the design, analysis or interpretation. The manuscript was then sent out to 420 reviewers from the journal’s database. Of the 221 reviewers who returned a review, the mean number of weaknesses identified was two, with 16% not identifying any weaknesses. Only 10% identified four or more weaknesses.

While peer review can be a challenge for a journal intending to publish high-quality science, for many OA journals, peer review is often simply absent. This is most apparent in a report by Ray (19). The author submitted eight manuscripts in response to ten email solicitations from suspect journals. The manuscripts submitted were four papers written by the author’s 13-year-old daughter or her 15-year-old friend. The papers were submitted, unaltered, to the journals. Nine of the ten journals responded to the submission, with eight accepting for publication. Six of the eight accepted the manuscript without any revisions. The one journal that rejected the manuscript indicated that the word count was too low, with the editor indicating that if the paper were expanded, they would reconsider it.

Literature Appraisal

As is apparent from the framework of EBM, the link between developing a clinical (or research) question and arriving at an answer to that question, is appraisal of the literature. The default position for many professionals is to simply read the abstracts (and sometimes simply the conclusion) of each paper and then move to the next one. This habit runs the great risk of being misled. The most common reasons for being misled are: the conclusion in the abstract may not be supported by the data in the paper; it was the wrong study type for the question the authors were trying to answer; or the quality of the actual study was weak.

Most medical research papers have a similar structure. This is often referred to as the IMRaD format. This

stands for Introduction, Methods, Results, Discussion. For purposes of this review, I will focus on the two sections that are most frequently not read; the Materials and Methods section, and the References.

Overview of Article Structure

Review papers, society position statements or opinion pieces may not follow the IMRaD format. The Introduction is usually one to two paragraphs long. The two most important items within the Introduction will be the background of why the study was undertaken, and a clear articulation of the actual research question. These items are critical for a true appreciation of the paper. If the Introduction does not include a clearly articulated research question (i.e., “We sought to demonstrate...”), it should be taken as a red flag. The most important section of the paper is the Materials and Methods. This is the section that most readers skip or simply scan. Many publishers will unfortunately make the font smaller and offset, as to give the impression that there is nothing important contained within. (I will go into more detail about the Materials and Methods section below). The Results section is the portion of the paper that will display the actual data findings of the study. In this section the reader should find the actual data displayed that the authors collected. Often, articles will have a “Table 1” that shows an overview (demographics, baseline characteristics) of the population studied. Another red flag is if the Results section does not report data that address the research question the authors proposed in

the Introduction section of the manuscript. The Discussion section of the paper should be a straightforward description of what the authors believe are the implications of the results of their study. The authors should relate what they believe to be the top “take home messages,” place their results in context of prior publications, suggest future directions of research and describe the limitations of their study. If the authors do not describe the limitation to their results, it should be taken as a red flag as well. Many authors will use the Discussion section to wax philosophically about issues tangential to the focus of the research project. While it may be useful for the authors to place their results in a larger framework, the further a Discussion section wanders from the research question and data presented, the more concerning it is that the manuscript is being used as a platform to promote the author’s personal perspective.

Introduction

As noted earlier, the Introduction of a research paper should include a clearly articulated research question. For research papers, the clinical question will orient the reader as to what to expect from the Materials and Methods section.

Clinical Questions

Table 1 displays the broad grouping of clinical questions. Questions of Therapy involve primarily decisions on management (e.g., which was the best med-

Table 1: Groupings of Clinical Questions; Adapted From (24)

Therapy	Testing the efficacy of drug treatments, surgical procedures, alternative methods of service delivery, or other interventions. Preferred study design is randomized controlled trial
Diagnosis	Demonstrating whether a new diagnostic test is valid (can we trust it?) and reliable (would we get the same results every time?). Preferred study design is cross sectional survey in which both the new test and the gold standard are performed
Screening	Demonstrating the value of tests which can be applied to large populations and which pick up disease at a presymptomatic stage. Preferred study design is cross sectional survey
Prognosis	Determining what is likely to happen to someone whose disease is picked up at an early stage. Preferred study design is longitudinal cohort study
Causation	Determining whether a putative harmful agent, such as environmental pollution, is related to the development of illness. Preferred study design is cohort or case-control study, depending on how rare the disease is, but case reports may also provide crucial information
Economic	To identify cost, cost-effectiveness, or cost-utility of a treatment, strategy, or service. Additionally, they can determine economic (or societal) burden of a disease, often accounting for cost of care and opportunity costs (i.e. lost wages, early mortality)

ication for a particular condition). The other clinical questions include questions of Prognosis (expected course of disease), Causation (do what degree does something cause disease), Diagnosis (what strategy is the best to identify condition; is a test valid), Screening (population strategies to identify disease) and Economic Evaluation (strategies which utilize fewest resources for given outcome). Each of these different types of clinical questions are answered using a different study design. Clinical practitioners are most familiar with research questions dealing with appropriate therapies (“Which treatment is the most appropriate for a particular disease?”) These clinical questions are best answered by an RCT (24); although other study designs may provide additional meaningful evidence. Questions of disease prognosis are also best answered with RCTs as well, but they often are unethical or infeasible. Absent an RCT, the best study design for a prognosis question is a cohort study of a similar population (25). If the prognosis question is about disease course (if picked up early vs. late), the preferred study design is a longitudinal cohort study (24). Questions about diagnosis are best answered using a cross-sectional study design in which a “gold standard” is used for comparison (24, 26). Screening tests are best evaluated using a cross-sectional survey in which the cohort or population receives both strategies (validation study) (24, 26).

Causation questions are best answered, depending upon the rarity of the condition, using a cohort or case-control study (24). As a matter of epistemology, how to “prove” causation remains debated. There exist no specific research criteria by which one can “prove” that an outcome has a specific cause; only demonstrations of a relationship between outcome and (proposed) cause. For clinical research the Bradford Hill Criteria (27, 28) serve as agreed upon guide for demonstrating causation. These criteria are listed in **Table 2**. Of these criteria, only one (temporality) is deemed as a requirement. The other criteria provide support for a causal relationship.

Materials and Methods

As noted earlier, the Materials and Methods section of a paper is the most important section. It should be the portion that is read first to help the reader decide if the paper is worth spending time on in the first place. If, from reading the methodology of the study, it is readily apparent that the study is the wrong design or is poorly done, then it would be better not to read the paper and run the risk of being distracted or misled.

Table 2: Bradford Hill Criteria for Causality; Adapted From (27)

1)	Strength	There is a strong association between the cause and the effect
2)	Consistency	There is a consistent association between cause and effect within comparable studies
3)	Specificity	There are a limited number of competing potential causes for the effect
4)	Temporal sequence*	The effect should always follow the cause
5)	Dose response	The cause and effect have a dose dependent relationship
6)	Biological plausibility	The relationship between the cause and the effect should be biologically reasonable; should not break the laws of physics
7)	Coherence	The association between the cause and effect should be consistent with other known biological causes; not an outlier
8)	Experimental evidence	The association between the cause and effect should be supported by experimental evidence
9)	Analogy	The mechanisms or processes in which the cause results in the effect should have other examples in nature

*The only required criterion

Study Designs

To gain a full understanding of the relative strength of a paper, the reader should have a working knowledge of the different study designs employed in clinical research. Each study design has its own strengths and weaknesses. A detailed description of study designs has been published earlier in this journal (29).

Case reports describe the clinical data on a specific patient (24, 30). The report typically focuses on a particular finding, feature or outcome that is noteworthy to the author. While case reports are common in the medical literature they often present clinical information which, often do not accurately represent the true nature of a disease or condition. Case reports typically describe a clinical oddity and should be viewed as hypothesis generating. When reading a case report, it is important to understand the context of case details. Questions to ask would include: Is this case reflective of similar cases in your experience? Is there something specific about this case that makes it an outlier?

Case series are a small number (typically fewer than ten) of collected case reports with a similar underlying feature (30). Similar to case reports, case series are hypothesis generating and need to be interpreted in context. How the series was accumulated very much influences the conclusions that could be made. Ideally, the subjects should be accumulated sequentially. If subjects were not collected sequentially they may have been subject to a surveillance bias which resulted in their being identified. An example of a sequentially collected series was reported by Love and colleagues (31). The primary purpose of the study was to systematically identify infant rib fracture patterns. All infants under 12 months of age had a full forensic anthropologist skeletal evaluation for a predefined 12-month period.

Cross-sectional surveys are reports of a particular question or finding in large number of subjects at a single point in time (24). These studies can provide valuable data on disease prevalence or norms and are often used in public health research. As cross-sectional studies often report on large samples of subjects,

conclusions about an individual subject are limited. This is referred to as the “Ecological Fallacy” (32).

Case-control studies are when subjects with a particular disease or finding (the “outcome”) are often matched to subjects without the disease or finding for comparison (the “exposure”) (24). The power of a case-control study is the ability to “look back” in time for subjects with a rare condition. The matching of the case and a control is an attempt to replicate the effect of randomization by balancing between the two groups any important “known” variables. By comparing otherwise similar subjects with and without a rare condition or finding, earlier risks (or “exposures”) may be identified. To assess the quality of a case-control study it is important for a paper to clearly identify on which characteristics/variables the cases were matched to the controls. Additionally, there may be some “unknown” variables which may account for differences found between the two groups that randomization would have negated. An example of a case-control study was reported by Price et al. (33). The researchers identified all children under ten years of age who died from blunt abdominal trauma over a 16-year period. These 33 children were then compared (but not matched) with children who died of natural causes who received cardiopulmonary respiration (CPR) (comparison group 1) and died of non-vehicular blunt abdominal trauma (comparison group 2) over the same period. The purpose of the study was to identify if CPR result in abdominal injuries which mimicked those seen in blunt abdominal trauma. As a result of this study, the authors concluded “*The likelihood of CPR-related primary abdominal trauma in child homicides is very low*” (33).

Cohort studies report on two (or more) groups of subjects with different risks (or “exposures”) who are followed for a period of time to assess outcomes (24). This study design is commonly used when randomization of the “exposure” would be unethical (smoking) or impossible (poverty) to perform. This design often requires many years for follow-up, but can provide very valuable data on disease causation or outcomes. Similar to case series studies, it is important for cohort studies to clearly identify how the cohort was col-

lected. If not clearly described, there may be hidden variables which account for reported findings. For example, caution needs to be exercised amongst the 35 infants and children identified by Matshes et al. who were accumulated from three different medical examiner offices over an undefined period of time (34). As inclusion criteria were not included (how each subject was identified) and the subjects were not sequentially included, conclusions need to be guarded, as the subjects may have been a biased, nonrepresentative cohort.

Randomized controlled trial (RCT) is the study design with which most physicians are familiar. In this design, two (or more) groups of subjects are randomly allocated to a “treatment” arm or a “control” arm (24). Both groups are then followed to assess the effects of the treatment. Randomized controlled trials can be single-blind (the subjects are unaware of which group they are in) or double-blind (neither the subject nor the investigator are aware of group allocation). The process of randomization allows for unknown important variables (e.g., age, gender, health) to be evenly balanced between the two groups. This balance allows the investigator to conclude that differences between the groups at the end of the study are due to the effects of the treatment. True RCTs are not a common study design in forensic pathology. Subjects can be sequentially collected and placed into “case” and “control” groups (e.g., traumatic, nontraumatic), but cannot be randomized into those groups. This is an echo of some of the criticisms of EBM noted earlier; RCTs are not applicable in many circumstances. Important quality measures of an RCT include true blinding of subjects (and, ideally, the investigator), true randomization (as opposed to haphazard), each group receiving the same measures, and outcome evaluations (ideally by blinded assessors).

Systematic reviews utilize a comprehensive literature search to answer a particular question. High quality systematic reviews begin with a focused and narrowly defined clinical question. The search protocol typically involves dozens of keyword combinations used to search within a number of different databases. The intention of the systematic review is be able to reflect

the sum total of the medical literature regarding the particular question (35). If studies within the systematic review are similar enough, their data can be combined into a meta-analysis. From within a systematic review, a meta-analysis takes the data from two or more studies and combines them in an effort to report a combined result, as if the studies were all part of a larger, single study. This statistical maneuver allows for a single overall estimated result from the body of literature analyzed. This single overall estimate is an attempt to report what the literature supports regarding a specific clinical question.

It is important to note that while systematic reviews and meta-analyses are viewed as study designs within the EBM framework, they are more appropriately identified as strategies for literature appraisal and interrogation. Alper and Haynes frame them as a lens through which a better appreciation of the underlying literature can be achieved (36).

Methodological Pitfalls

Bias

One of the primary strength of the systematic critical appraisal is an attempt to overcome bias. Bias, in clinical research, is systematic error that can artificially distort the data (37, 38). Bias often is a function of poor study design or implementation. An example of bias would be identifying a cohort of patients for a study in an inconsistent or arbitrary manner (selection bias). This cohort may have characteristics which are artificially present and not representative of the real population. **Table 3** lists a number of other biases that can influence studies. Common methods to reduce bias in research are subject randomization, consecutive recruitment of subjects, prospective study design, and investigator blinding (38).

Circular Reasoning

This is one of the most critical aspects of research design that requires careful attention when reading the Materials and Methods section. Circular reasoning within clinical research is, broadly, when the outcome

is part of the inclusion criteria (39). This is often referred to as “begging the question.” This is a common issue in many medical and forensic studies when there exists no “gold standard” for the outcome (e.g., sudden infant death syndrome, child abuse, mental illness, migraines). When outcomes or inclusion parameters are determined by consensus, scoring tools, or composite criteria, it becomes more difficult to ensure a clear distinction between the two. This alone would not indicate that the paper is not valuable, but the firm or broad conclusions should be tempered and the results should be placed into context of other published literature.

Discussion

The Discussion section is where the authors describe the implications of their study. The authors should place their results in context of practice or other published data, should highlight any practice implications, and should identify potential next steps that their study leads to. It is also important that authors include a description of the limitations that their study has. Be very cautious of a paper that is not explicit in its limitations; all studies have them.

Rhetorical Pitfalls

Within the Discussion section, authors will try to frame their argument for conclusions based upon their results. Like all arguments, the reader should be mindful of how the authors try to make their case for conclusions they are promoting. There are a number of rhetorical maneuvers that authors may intentionally or unintentionally make that pose pitfalls to the reader. I will highlight some of the most common within the medical and forensic literature.

Association is not Causation

As noted earlier, the clinical research has limited ability to “prove” causation. The Bradford Hill criteria (27) provide a framework to interpret potential causal relationships. The pitfall occurs with over generous interpretation of a relationship between two events, and assigning causation simply because one event occurs after the other (*cum hoc ergo propter hoc*, or “after this, therefore because of this”). There may be a causal relationship, but they both may be related to a third event (which was unmeasured) and simply associated with each other through the third event. For

Table 3: Potential Biases in Research and Clinical Practice; Adapted From (38)

Bias	Description
Selection bias	The subjects, interventions or procedures are chosen in a non-random fashion which may affect the results
Sample bias	The group chosen to study does not represent the population of interest
Loss-to-follow-up bias	Subjects in a study are followed-up in an unequal manner
Disease spectrum bias	A limited form of the condition is studied; not representative of the true spectrum of the disease
Referral bias	Local practices or conditions affect patient referrals or procedures
Self-selection bias	The subjects who self-select are different from those who do not
Recall bias	Subject historical recall can be incomplete to “unremarkable” but important items
Interviewer bias	An interviewer may “frame” or “coach” the information from the subject
Verification bias	Testing occurs to a subset of subjects who “screen positive” and not to a full population
Response bias	Data are missing in a non-random manner; negative results are not evaluated as thoroughly
Reviewer bias	The person collecting the data is not blinded to disease or condition
Test review bias	In retrospective studies; when the diagnosis is known at the beginning of the study
Imperfect-standard bias	The diagnostic standard is subjective
Confounding	The role other “unknown” variables play in the data; within the subject, disease, or study design

example, people who frequent pool halls have a higher rate of lung and esophageal cancers. The pool halls are not responsible for the increased cancer rates. It is due to the fact that people who spend time in pool halls drink alcohol and smoke more than those who do not. While there is an association between pool hall and lung and esophageal cancers, the increased cancer rate is due to the drinking and smoking and not the pool hall.

Hasty Generalizations

Hasty generalizations, as the name suggests, is when the breadth and scope of conclusions of a research project, far outstrip the data the research generated. This is most clearly seen by the over dependence on using case reports to support clinical or forensic decisions. As noted earlier, case reports are by definition, outliers. To take conclusions generated as a result of a single case report and then generalize or extrapolate them runs the risk of being truly misled. A clear example would be the case report of a 15-year-old girl who survived untreated rabies (a uniformly fatal condition) (40). To then extrapolate that rabies is now a survivable disease would be the incorrect conclusion. The authors of the report rightly concluded “*Survival of this single patient does not change the overwhelming statistics on rabies, which has the highest case fatality ratio of any infectious disease.*”

Argument from Ignorance (*argumentum ad ignorantium*)

The logical fallacy pitfall makes the argument that because something has not been shown to be true, it must be false. Likewise, the reverse is also a logical fallacy (because something has not been shown to be false, it must be true) (39). There are many meaningful truths that we believe to be true (or false) that don’t require research to demonstrate. These are usually dangerous, or unethical (or simply unfounded) to critically study (e.g., jumping out of an airplane without a parachute is dangerous, antibiotics should be used for bacterial meningitis, surgical outcomes are improved if scrubs are blue).

Straw Man

Most people are familiar with the phenomenon of the straw man argument. This is mostly popular in the political realm, and describes when an opposing position is misrepresented in such a manner that it becomes clearly inferior. This misrepresented view is then readily refuted, claiming victory (41). In a manuscript, this would appear within the discussion section when presenting an opposing view. This opposing view would be skewed in a way to make it easy for the reader to find it not credible. This can be explored by reading the references which are used to represent to opposing view and ensure that it is being presented accurately.

References

The references are the listing of citations that the author has used to support the points made in the manuscript. This listing is conventionally at the end of the manuscript. There are two common conventions of notations within the article. The most common in medical and scientific publications is sequential numbers (either superscript or parenthetical) at the end of the sentence. Within the social sciences, it is common to have citations (the last name of the first author and year of publication) in a parenthesis embedded within the manuscript. The references are then listed at the end of the article in alphabetical order by last name of first author. It is quite important for the reader to scrutinize the references which an author uses. There are two common issues that can be regularly identified by a close reading of the references.

The first is chain citation. This is when, after being repeatedly cited, the original fact is no longer accurately reported. This is akin to the childhood game of telephone; with a phrase being repeated into the ear of a neighbor, and at the end of the circle, the original phrase has transformed in the retelling. An excellent example of this is reported by Casaulta et al. regarding the oft described phenomenon in children with G-6-PD deficiency having an abnormal sweat chloride test (42). The authors report that this is a truism often taught, but the authors were unable to identify the source of data for the original tables. This alleged

false-positive sweat testing results are regularly included in textbook tables in apparent cut-and-paste fashion after being hypothesized at a conference in 1975, despite data never having been reported in published literature.

The second common issue uncovered with sad regularity is the citation used by the author does not actually support the point in the manuscript. This is often the result of sloppy attribution. The authors may believe that a particular paper supports the point that is being made, but they do not actually read the reference to confirm that it actually does. While this will increase the effort required to read an article, reading the references used to support specifically critical points within a manuscript will regularly reveal how frequently there is asynchrony between the reference and the way it is portrayed in the paper.

CONCLUSION

Critical literature appraisal is now akin to hand-to-hand combat. No longer can we read articles secure in the comfort that we are not being led astray. We must now be active participants in reading the medical literature. There are four main recommendations for the reader. First, be skeptical of everything that you read. We must now evaluate the provenance of the journal by checking it against a "Black List" of suspicious publishers and journals. Second, read the Materials and Methods section of the paper. We often skim or skip it, but we can no longer. While many journals will print the Materials and Methods section offset and in a smaller font, giving the impression that it contains nothing meaningful, it should be the first thing we read. This section will tell us if we should read the paper or not. Third, read the references. Critical information can be found within the articles that the author felt were important to their paper. It will also be a check of the quality of the scholarship of the author; confirming that the references support what the author says they support. Fourth, remember that peer review is not a guarantee that the paper is of high-quality. While peer review should be a minimum criteria for a manuscript, in many cases, it may only mean that someone else also read the paper.

Ultimately it is up to us to read the medical literature carefully, scrutinize it thoughtfully, and place the data in the context of other published data and our clinical experiences.

REFERENCES

- 1) Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA*. 1992 Nov 4; 268(17):2420-5. PMID: 1404801. <https://doi.org/10.1001/jama.268.17.2420>.
- 2) Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996 Jan 13; 312(7023):71-2. PMID: 8555924. PMCID: PMC2349778. <https://doi.org/10.1136/bmj.312.7023.71>.
- 3) Guyatt G, Rennie D, Hayward R. Users' guides to the medical literature: a manual for evidence-based clinical practice. Chicago: AMA press; 2002. 706 p.
- 4) Straus SE, Richardson WS, Glasziou P, Haynes RB. Evidence-based medicine: how to practice and teach EBM. New York: Elsevier/Churchill Livingstone; 2005. 299 p.
- 5) Sackett DL. Evidence-based medicine. *Semin Perinatol*. 1997 Feb; 21(1):3-5. PMID: 9190027.
- 6) Evidence-based medicine, in its place. *Lancet*. 1995 Sep 23; 346(8978):785. PMID: 7674736. [https://doi.org/10.1016/s0140-6736\(95\)91610-5](https://doi.org/10.1016/s0140-6736(95)91610-5).
- 7) White KL, Fowler PBS, Bradley F, et al. *Lancet*. 1995 Sep 23; 346(8978):837-40. PMID: 7674753. [http://dx.doi.org/10.1016/S0140-6736\(95\)91651-2](http://dx.doi.org/10.1016/S0140-6736(95)91651-2).
- 8) Grahame-Smith D. Evidence based medicine: Socratic dissent. *BMJ*. 1995 Apr 29; 310(6987):1126-7. PMID: 7742683. PMCID: PMC2549506. <https://doi.org/10.1136/bmj.310.6987.1126>.
- 9) Thomas SJ. Does evidence-based health care have room for the self? *J Eval Clin Pract*. 2016 Aug; 22(4):502-8. PMID: 27237731. <https://dx.doi.org/10.1111/jep.12563>.
- 10) van Baalen S, Boon M. An epistemological shift: from evidence-based medicine to epistemological responsibility. *J Eval Clin Pract*. 2015 Jun; 21(3):433-9. PMID: 25394168. <https://dx.doi.org/10.1111/jep.12282>.
- 11) Seglen PO. Why the impact factor of journals should not be used for evaluating research. *BMJ*. 1997 Feb 15; 314(7079):498-502. PMID: 9056804. PMCID: PMC2126010. <https://doi.org/10.1136/bmj.314.7079.497>.
- 12) The impact factor game. It is time to find a better way to assess the scientific literature. *PLoS Med*. 2006 Jun; 3(6):e291. PMID: 16749869; PMCID: PMC1475651. <https://dx.doi.org/10.1371/journal.pmed.0030291>.
- 13) Kurmis AP. Understanding the limitations of the journal impact factor. *J Bone Joint Surg Am*. 2003 Dec; 85-A(12):2449-54. PMID: 14668520. <https://doi.org/10.2106/00004623-200312000-00028>.
- 14) Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 1953 Apr 25; 171(4356):737-8. PMID: 13054692. <https://doi.org/10.1038/1038/171737a0>.
- 15) Wakefield AJ, Murch SH, Anthony A, et al. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet*. 1998 Feb 28; 351(9103):637-41. PMID: 9500320. [https://doi.org/10.1016/s0140-6736\(97\)11096-0](https://doi.org/10.1016/s0140-6736(97)11096-0). Retraction in: *Lancet*. 2010 Feb 6; 375(9713):445. PMID: 20137807. [https://dx.doi.org/10.1016/S0140-6736\(10\)60175-4](https://dx.doi.org/10.1016/S0140-6736(10)60175-4).

- 16) Fraser AG, Dunstan FD. On the impossibility of being expert. *BMJ*. 2010 Dec 14; 341:c6815. PMID: 21156739. <https://doi.org/10.1136/bmj.c6815>.
- 17) Wager E, Williams P. Why and how do journals retract articles? An analysis of Medline retractions 1988-2008. *J Med Ethics*. 2011 Sep; 37(9):567-70. PMID: 21486985. <https://dx.doi.org/10.1136/jme.2010.040964>.
- 18) Beninger PG, Beall J, Shumway SE. Debasing the currency of science: the growing menace of predatory open access journals. *J Shellfish Res*. 2016; 35(1):1-5. <http://dx.doi.org/10.2983/035.035.0101>.
- 19) Ray M. An expanded approach to evaluating open access journals. *J Sch Publ*. 2016; 47(4):307-27. <https://doi.org/10.3138/jsp.47.4.307>.
- 20) Byard RW. The forensic implications of predatory publishing. *Forensic Sci Med Pathol*. 2016 Apr 2. [Epub ahead of print]. PMID: 27038941. <https://dx.doi.org/10.1007/s12024-016-9771-3>.
- 21) Bohannon J. Who's afraid of peer review? *Science*. 2013 Oct 4; 342(6154):60-5. PMID: 24092725. <https://dx.doi.org/10.1126/science.342.6154.60>.
- 22) PubMed Bethesda (MD): U.S. National Library of Medicine, National Institutes of Health; 2016 [cited 2016 Oct 5]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed>.
- 23) Godlee F, Gale CR, Martyn CN. Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial. *JAMA*. 1998 Jul 15; 280(3):237-40. PMID: 9676667. <https://doi.org/10.1001/jama.280.3.237>.
- 24) Greenhalgh T. How to read a paper. Getting your bearings (deciding what the paper is about). *BMJ*. 1997 Jul 26; 315(7102):243-6. PMID: 9253275. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2127173/>.
- 25) Geddes J. Answering clinical questions about prognosis. *Evid Based Ment Health*. 2000; 3(4):100-1. <https://doi.org/10.1136/ebmh.3.4.100>.
- 26) Greenhalgh T. How to read a paper: Papers that report diagnostic or screening tests. *BMJ*. 1997 Aug 30; 315(7107):540-3. PMID: 9329312. PMCID: PMC2127365. <https://doi.org/10.1136/bmj.315.7107.540>.
- 27) Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965 May;58:295-300. PMID: 14283879. PMCID: PMC1898525. <https://doi.org/10.1177/0141076814562718>.
- 28) Schünemann H, Hill S, Guyatt G, Akl EA, Ahmed F. The GRADE approach and Bradford Hill's criteria for causation. *J Epidemiol Community Health*. 2011 May;65(5):392-5. PMID: 20947872. <https://dx.doi.org/10.1136/jech.2010.119933>.
- 29) Pinckard JK. Critical appraisal of scientific literature. *Acad Forensic Pathol*. 2011 Nov; 1(3):302-9.
- 30) Vandenbroucke JP. In defense of case reports and case series. *Ann Intern Med*. 2001 Feb 20; 134(4):330-4. PMID: 11182844. <https://doi.org/10.7326/0003-4819-134-4-200102200-00017>.
- 31) Love JC, Derrick SM, Wiersema JM, et al. Novel classification system of rib fractures observed in infants. *J Forensic Sci*. 2013 Mar; 58(2):330-5. PMID: 23406328. <https://dx.doi.org/10.1111/1556-4029.12054>.
- 32) Piantadosi S, Byar DP, Green SB. The ecological fallacy. *Am J Epidemiol*. 1988 May; 127(5):893-904. PMID: 3282433.
- 33) Price EA, Rush LR, Perper JA, Bell MD. Cardiopulmonary resuscitation-related injuries and homicidal blunt abdominal trauma in children. *Am J Forensic Med Pathol*. 2000 Dec; 21(4):307-10. PMID: 11111786. <https://doi.org/10.1097/00000433-200012000-00001>.
- 34) Matshes EW, Evans RM, Pinckard JK, et al. Shaken infants die of neck trauma, not of brain trauma. *Acad Forensic Pathol*. 2011 Jul; 1(1):82-91.
- 35) Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med*. 1997 Mar 1; 126(5):376-80. PMID: 9054282. <https://doi.org/10.7326/0003-4819-126-5-199703010-00006>.
- 36) Alper BS, Haynes RB. EBHC pyramid 5.0 for accessing preappraised evidence and guidance. *Evid Based Med*. 2016 Aug; 21(4):123-5. PMID: 27325531. <https://dx.doi.org/10.1136/ebmed-2016-110447>.
- 37) Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. *Plast Reconstr Surg*. 2010 Aug; 126(2):619-25. PMID: 20679844. PMCID: PMC2917255. <https://dx.doi.org/10.1097/PRS.0b013e3181de24bc>.
- 38) Sica GT. Bias in research studies. *Radiology*. 2006 Mar; 238(3):780-9. PMID: 16505391. <https://dx.doi.org/10.1148/radiol.2383041109>.
- 39) Hahn U, Oaksford M. The rationality of informal argumentation: a Bayesian approach to reasoning fallacies. *Psychol Rev*. 2007 Jul; 114(3):704-32. PMID: 17638503. <https://dx.doi.org/10.1037/0033-295X.114.3.704>.
- 40) Willoughby RE Jr., Tieves KS, Hoffman GM, et al. Survival after treatment of rabies with induction of coma. *N Engl J Med*. 2005 Jun 16; 352(24):2508-14. PMID: 15958806. <https://dx.doi.org/10.1056/NEJMoa050382>.
- 41) Talisse R, Aikin SF. Two forms of the straw man. *Argumentation*. 2006; 20(3):345-52. <https://doi.org/10.1007/s10503-006-9017-8>.
- 42) Casaulta C, Stirnimann A, Schoeni MH, Barben J. Sweat test in patients with glucose-6-phosphate-1-dehydrogenase deficiency. *Arch Dis Child*. 2008 Oct; 93(10):878-9. PMID: 18456694. <https://dx.doi.org/10.1136/adc.2007.132688>.