ELSEVIER

Journal of Clinical Epidemiology 66 (2013) 361−366

Journal of Clinical Epidemiology

REVIEW ARTICLES

Observational studies often make clinical practice recommendations: an empirical evaluation of authors' attitudes

Vinay Prasad[a], Joel Jorgenson[b], John P.A. Ioannidis[c,d], Adam Cifu[e,*]

[a]*Medical Oncology Branch, National Cancer Institute, National Institutes of Health, 10 Center Dr., 10/12N226, Bethesda, MD 20892, USA*
[b]*Feinberg School of Medicine, Northwestern University, Chicago, IL, USA*
[c]*Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Room X306, 1265 Welch Rd., Stanford, CA, USA*
[d]*Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA, USA*
[e]*Department of Medicine, Pritzker School of Medicine, University of Chicago Medical Center, 5841 South Maryland Avenue, Chicago, IL 60637, USA*

Accepted 25 November 2012; Published online 4 February 2013

Abstract

Objectives: Although observational studies provide useful descriptive and correlative information, their role in the evaluation of medical interventions remains contentious. There has been no systematic evaluation of authors' attitudes toward their own nonrandomized studies and how often they recommend specific medical practices.

Study Design and Setting: We reviewed all original articles of nonrandomized studies published in 2010 in *New England Journal of Medicine*, *Lancet*, *Journal of the American Medical Association*, and *Annals of Internal Medicine*. We classified articles based on whether authors recommend a medical practice and whether they state that a randomized trial is needed to support their recommendation. We also examined the types of logical extrapolations used by authors who did advance recommendations.

Results: Of the 631 original articles published in 2010, 298 (47%) articles were eligible observational studies. In 167 (56%) of 298 studies, authors recommended a medical practice based on their results. Only 24 (14%) of 167 studies stated that a randomized controlled trial (RCT) should be done to validate the recommendation, whereas the other 143 articles made a total of 149 logical extrapolations to recommend specific medical practices. Recommendations without a call for a randomized trial were most common in studies of modifiable factors (59%), but they were also common in studies reporting incidence or prevalence (51%), studies examining novel tests (41%), and association studies of nonmodifiable factors (32%).

Conclusion: The authors of observational studies often extrapolate their results to make recommendations concerning a medical practice, typically without first calling for a RCT. Published by Elsevier Inc.

Keywords: Epidemiology; Clinical trials; Observational studies; Randomized trials; Hierarchy of research design; Reversal

1. Introduction

Observational studies play an important role in advancing medical knowledge. They yield crucial data on incidence, prevalence, correlation, association, prognosis, and natural history. Their role, however, in answering questions regarding medical practices—for example, the use of treatments and diagnostic and screening tests—has long been a contentious issue. One early empirical evaluation compared the results of historical studies with those of historical controls vs. randomized controlled trials (RCTs) [1]. For six different therapies tested in 50 RCTs and 56 studies with historical controls, the authors found that a particular agent was considered effective in 79% of studies with historical controls but only in 20% of RCTs. In 2000, two high-profile empirical evaluations [2,3] found remarkable agreement between the two types of design. These conclusions met with criticism [4,5], and a greater proportion of disagreement was found in the largest empirical evaluation [6] (of 45 topics and 408 studies), with differences in the effect size exceeding 50% seen in 62% of the topics. Other empirical evaluations have found that five of the six most cited observational studies were refuted or found to have exaggerated results when tested in RCTs [7]. The discrepancy rate between observational studies and randomized trials may vary according to topic, with greater discordance in some fields such as nutrition and cancer and better agreement in other types of questions such as appraisal of harms of medical interventions [8−11].

Conflict of interest statement: The authors declare no conflicts of interest.

* Corresponding author. Tel.: (773) 702-6840; fax: (773) 834-7492.
E-mail address: acifu@medicine.bsd.uchicago.edu (A. Cifu).

0895-4356/$ - see front matter Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.jclinepi.2012.11.005

What is new?

- The majority of authors (56%) of observational studies in high impact journals make medical practice recommendations. These extrapolations may not logically follow from the authors' own research, and may instead be best supported by prospective, randomized studies. Nevertheless, only a minority (14%) of these authors call for a randomized controlled trial to support their recommendation prior to implementation.

Biases may exist in both RCTs and observational studies [12], and both types of studies are useful. However, making inferences about medical treatments and management based on observational studies alone may be precarious. Even large well-done observational studies may be frequently wrong [13], and the association does not prove causation. Although these limitations are recognized, it is unknown whether the authors of observational studies acknowledge them and abstain from making recommendations regarding medical practice.

It is also worth noting that not all observational studies address medical treatments. Some observational studies may provide estimates of incidence and prevalence of a disease. Other types of observational studies may address the performance of a diagnostic or screening test or demonstrate that some agent is a risk factor for (or protects against) disease. Yet, even in these cases, authors may recommend a clinical practice in their article. Such recommendations often do not logically follow from the data they have presented. It would thus be interesting to systematically appraise the authors' attitudes toward their own nonrandomized studies.

Here, we sought to investigate the authors' attitudes toward observational studies. We examined all original articles from four major general medicine journals in 2010. We aimed to evaluate how often authors state that their work supports a stance toward a medical practice, and if so, whether they stated that a randomized trial would be necessary to support their recommendation.

2. Methods

2.1. Eligible studies

We examined all original articles published in one calendar year in *New England Journal of Medicine (NEJM)*, *Journal of the American Medical Association (JAMA)*, *Lancet*, and *Annals of Internal Medicine*, the four general medicine journals with highest impact factor according to *Journal Citation Reports* 2010 edition. Articles considered were listed under the heading ''Articles'' in *Annals of Internal Medicine*, ''Original Contribution'' in *JAMA*, ''Original Article'' in *NEJM*, and ''Original Research Article'' in *Lancet*. We chose 2010 as it represents the last complete year at the time we started our investigation. Articles were reviewed independently by two reviewers (V.P. and J.J.). We excluded all RCTs, meta-analyses, systematic reviews, decision and cost-effectiveness analyses, studies using nonhuman subjects, and studies whose main data were derived from modeling. We included only case series with $N > 5$. Both retrospective and prospective and controlled and uncontrolled observational studies were included. This review was performed in duplicate, with strong intrarater agreement (Cohen kappa, 0.97).

2.2. Data extraction

For each included study, we assessed whether the authors recommended some course of medical practice, and if so, what their recommendation(s) is (are). Medical practice was defined broadly and included screening, diagnostic, and treatment-guiding tests; medications; interventions; other therapeutics; behavioral or counseling recommendations; changes to hospital or systems structure; or, broadly, any activity that might be performed by some member of a health care team. Specific recommendation statements were evaluated independently by two investigators with strong agreement (Cohen kappa, 0.96) and recorded verbatim. Furthermore, we recorded whether the authors' stated that an RCT was needed to support the proposed recommendation. This too was performed in duplicate (Cohen kappa, 0.92).

2.3. Classification

One reviewer (V.P.) classified each included study as one of four types: incidence/prevalence studies (those that reported the incidence or prevalence of some medical illness or practice), treatment association studies (those that examined associations or correlations between modifiable factors and outcomes), testing studies (those that examined diagnostic, screening, or stratification tests), and all other association studies involving nonmodifiable factors. The distinction between treatment association and other association studies is that the former examines associations between some factor or practice, which is under the control of health care personnel or patients, and some outcome (for instance, the administration of a therapy with mortality), whereas ''other associations'' examine the association between two phenomena outside the control of health care personnel or patients (for instance, the relationship between one illness and another).

When nonrandomized studies made recommendations for medical practices, four kinds of logical leaps (extrapolations) were noted. In incidence/prevalence studies, authors may have used their article to argue that some specific remedy should be performed regarding the illness or practice being studied. However, simply because something is prevalent does not mean it is alterable, and furthermore,

incidence/prevalence studies do not demonstrate that the specific recommendations would improve the problem. We named this extrapolation "Problems demand solutions." In treatment association studies, authors who found that some factor or practice is associated with a better or worse outcome may have concluded that we should try to influence this factor or adopt (or avoid) the respective practice. This claim is tenuous because correlation in an observational study does not imply causation. We named this extrapolation "Correlation implies causation." In testing studies, authors who showed that some test allows us to identify subgroups of patients who have a different disease status or experience a different, better, or worse outcome may then have argued that we should incorporate that test into clinical practice. This is problematic because discriminatory ability does not necessarily mean that the routine use of a test would ultimately allow for better outcomes [14]. We call this extrapolation "Discrimination implies utility." Finally, in any type of observational study, authors may have used the discussion or abstract section to make other practice recommendations that were not related to their investigation, besides pertaining to the same subject matter. We call this extrapolation "Other Assertions."

Besides descriptive statistics, comparisons across groups were performed using the chi-square and Fisher exact tests, as appropriate. *P*-values are two-tailed. A *P*-value of $<$ 0.025 was deemed significant.

3. Results

Overall, 631 original articles were published in the four journals in 2010. Of those, 298 (47%) articles were eligible observational studies. In 167 (56%) of the 298 studies, authors endorsed a recommendation regarding a medical practice. Of the 167 articles that make a recommendation regarding a medical practice, only 24 articles (14%) stated that an RCT should first be performed, whereas the other 143 (86%) articles made a recommendation without such a call. Fig. 1 shows this breakdown of articles graphically. Authors of articles who advanced recommendations called for an RCT at the same rate as authors of articles who did not (14% vs. 18%, $P = 0.52$). Extrapolations without calls for RCT were made by a substantial proportion of articles in all the four journals (ranging from 40% to 65% across journals).

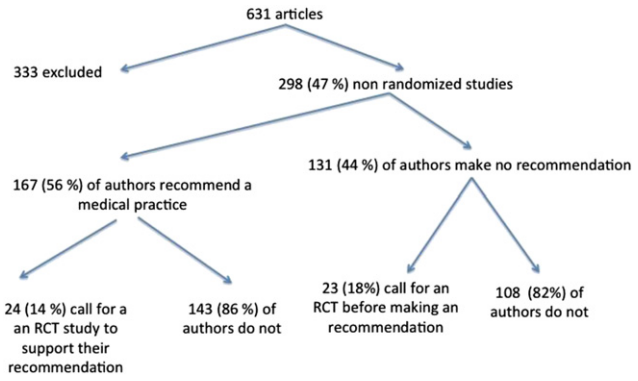

Fig. 1. A breakdown of examined articles. RCT, randomized controlled trial.

Of the 298 articles, 108 (36%) addressed incidence/prevalence, 82 (28%) were treatment associations (studies of modifiable factors), 59 (20%) addressed testing, and 49 (16%) pertained to other associations (nonmodifiable factors). The proportions of articles that made recommendations were 56%, 66%, 53%, and 43% ($P < 0.073$) for each category, respectively. The respective proportions that made recommendations without calling for RCTs were 51%, 59%, 42%, and 31% ($P < 0.013$). Thus, recommendations without calling for an RCT were most frequent when associations with modifiable factors were evaluated and least frequent when nonmodifiable factors were studied (Table 1).

Fig. 2 depicts the percentage of articles making each type of extrapolation among articles that did not call for RCT. Incidence/prevalence articles proposed specific solutions in 41 (28%) articles, made other assertions in 11 (10%) articles, and did both in another 3 (3%) articles. Treatment studies recommended (or cautioned against) the use of that treatment in 44 (54%) articles, made other assertions in 3 (4%) articles, and did both in another 1 (1%) article. Testing studies recommended the use of that test in 22 (37%) articles, made other assertions in 1 (2%) article, and did both in another 2 (3%) articles. Among other association studies, 15 articles made various assertions (31%).

Appendix 1 (available on the journal's website at www.jclinepi.com) depicts the examples of each of the four types of extrapolation and offers an explanation as to why the authors' recommendations are not necessarily supported by their study. For instance, one incidence/prevalence study showed that patients with advanced malignancy were frequently undergoing screening for other cancers [15]. Such a practice is clearly wasteful, but the authors recommend specifically that electronic medical records should flag these patients' charts. This recommendation is not supported by their article and may or may not achieve the desired outcome. Physicians may see such a reminder, workflow may be interrupted, but the unnecessary screening test may nevertheless be ordered. Empirical studies of electronic medical records support equipoise, noting that new protocols may disrupt care, while failing to alter outcomes [16,17].

One association study involving a modifiable factor linked the consumption of fructose with gout in women and then advised a decrease in fructose consumption as an effective intervention [18]. A testing/screening study identified rare chromosomal deletions in attention-deficit hyperactivity disorder (ADHD), through genome-wide analysis [19]. It then recommended routine referral to geneticists to screen for such mutations among children with ADHD. Finally, a article showed that hospitalization is

Table 1. Stance of observational studies toward making recommendations and making recommendations without even calling for an RCT

Types of studies	Number of articles	Number (%) making a recommendation	Number (%) making a recommendation without calling for an RCT
Incidence/prevalence	108	61 (56)	55 (51)
Treatment association (modifiable factors)	82	54 (66)	48 (59)
Testing	59	31 (53)	25 (42)
Other associations (nonmodifiable factors)	49	21 (43)	15 (31)
Total	298	167	143

Abbreviation: RCT, randomized controlled trial.

associated with dramatic functional decline and death among older persons [20]. The authors use this observation to make several recommendations for clinical practice (see Appendix 1, available on the journal's website at www.jclinepi.com).

4. Discussion

More than half of the nonrandomized studies published recently in four top impact medical journals make recommendations (56%) regarding a stance toward medical practice, and only a small fraction of those (14%) call for RCTs to support their endorsements. Extrapolations to recommendations are made in all types of observational studies. If anything, extrapolations without a call for an RCT are most frequent when associations with modifiable factors are studied. Recommendations without calling for RCTs were common in all four journals that we examined.

There exist ample arguments for and against the value of observational studies in informing and guiding treatment decisions [2–5]. Our study is relevant to those on both sides of the debate. Even if we believe in the accuracy of observational studies and their ability to predict treatment effects, this would only justify at best the logical leaps we refer to as Correlation implies causation. Our empirical evaluation shows however that most extrapolations are not of this kind (63%). Authors are eager to extrapolate to recommendations in many incidence/prevalence studies, testing studies, and studies of correlations with nonmodifiable factors, in which the study designs are typically unsuitable to inform recommendations by themselves.

It is also worth noting that many clinical groups take a harder stance and consider observational studies to be a poor basis for treatment recommendations. The Grading of Recommendations Assessment, Development and Evaluation Working Group designates four tiers for quality of evidence (high, moderate, low, and very low) [21].

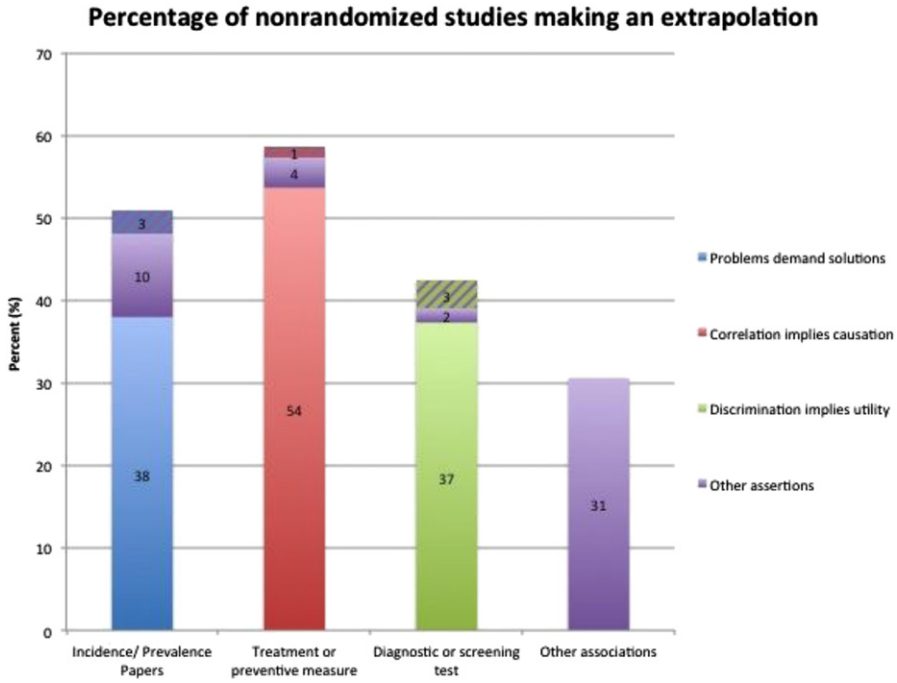

Fig. 2. Percentage of articles making extrapolations by the type of article. Striped bars (of two interlaced colors) indicate that those publications made two types of extrapolation (please refer the online version of the figure for colors).

Observational studies, without exceptional strengths or deep limitations, constitute low-grade evidence according to this framework [21]. As such, the extrapolation we call Correlation implies causation is likely also tenuous.

In many cases, we are very sympathetic to the recommendations of the authors. For example, efforts to improve the speed of detection of malaria in poor regions [22], combat violence against women [23], and reduce childhood obesity [24] are some of the least controversial recommendations we found. However, even in these cases, the authors' work and the proposed recommendations are probably disconnected. The authors show convincingly that these are major problems, but this realization does not inform which of the many possible strategies for amelioration is best.

It is common to see new studies contradict previous adopted standards of care [25,26]. Even the results of highly cited studies can be refuted [7], and the replication rate tends to be low for claims made from observational designs [7]. We have previously noted that the most common correlate for reversal of standards of care was the original adoption of a practice based on nonrandomized evidence alone [27]. The studies examined here offer many recommendations that may be precarious or erroneous. If adopted, such practices may need to be reversed in the future after having been detrimental to health, health finances, and the reputation of medical science.

There are several limitations to our study. First, the publication patterns of high-impact journals may not represent those of other lesser impact factor journals [28]. For example, in high-impact journals, the authors may feel more compelled to show that their work leads to important actions to justify publication in these venues. Recommendations made in such journals are likely to be influential given the prestige of the accompanying name branding. However, high-impact journals may also provide more critical review. For instance, *JAMA* [29] asks authors to avoid, ''speculation and overgeneralization'' in their discussion sections. Future studies may evaluate the stance of authors in observational studies published in specialty journals.

Second, as with any study of this kind, which requires some subjective judgment, it is possible that others will interpret any one study, or the authors' recommendations, differently than we did. We did however perform our analysis with two independent reviewers, duplicate data extraction, and high interrater reliability. Thus, we feel confident that the average reader would agree with our interpretation of authors' statements.

Third, for some types of medical practices, RCTs are almost never performed, but the practices are nevertheless adopted. In particular, the adoption of diagnostic and prognostic tests has often relied on nonrandomized evidence. There is substantial evidence that the literature regarding such novel tests is overstated [30], and other empirical evaluations have also noted the propensity of authors to extrapolate toward clinical applications in the presence of scant evidence [31]. As we become inundated with proposed tests, it is likely that we will require more robust evidence to identify the tests, which are worth adopting in clinical practice [14]. An individual test may be excellent at differentiating two populations of patients, but this discrimination may not lead to improved outcomes.

Fourth, we did not assess whether independent randomized data exist in support of the authors' recommendations. This would have required a close examination and systematic review of the literature on hundreds of topics. As an alternative, one might have evaluated the cited articles in each study. However, cited articles may offer a partial view of the evidence, typically supporting disproportionately the authors' claims, although sometimes they also propagate clearly erroneous conclusions [32,33].

In conclusion, our empirical evaluation shows that linking observational results to recommendations regarding medical practice is currently very common in highly influential journals. Such recommendations frequently represent logical leaps. As such, if they are correct, they may accelerate the translation of research but, if they are wrong, they may cause considerable harm.

Acknowledgments

V.P. had full access to all the data in the study and takes responsibility for the integrity of the data and accuracy of the data analysis.

References

[1] Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. Am J Med 1982;72:233−40.
[2] Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. New Engl J Med 2000;342:1887−92.
[3] Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med 2000;342:1878−86.
[4] Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? N Engl J Med 2000;342:1907−9.
[5] Ioannidis JP, Haidich AB, Lau J. Any casualties in the clash of randomised and observational evidence? BMJ 2001;322:879−80.
[6] Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001;286:821−30.
[7] Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. JAMA 2005;294:218−28.
[8] Shikata S, Nakayama T, Noguchi Y, Taji Y, Yamagishi H. Comparison of effects in randomized controlled trials with observational studies in digestive surgery. Ann Surg 2006;244:668−76.
[9] 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;38:1586−91.
[10] Boffetta P, McLaughlin JK, La Vecchia C, Tarone RE, Lipworth L, Blot WJ. False-positive results in cancer epidemiology: a plea for epistemological modesty. J Natl Cancer Inst 2008;100:988−95.
[11] Meyskens FL Jr, Szabo E. Diet and cancer: the disconnect between epidemiology and randomized clinical trials. Cancer Epidemiol Biomarkers Prev 2005;14:1366−9.

[12] Gluud LL. Bias in clinical intervention research. Am J Epidemiol 2006;163:493−501.
[13] Ioannidis JP. Why most published research findings are false. PLoS Med 2005;2:e124.
[14] Ioannidis JP, Tzoulaki I. What makes a good predictor?: the evidence applied to coronary artery calcium score. JAMA 2010;303:1646−7.
[15] Sima CS, Panageas KS, Schrag D. Cancer screening among patients with advanced cancer. JAMA 2010;304:1584−91.
[16] Sheikh A, Cornford T, Barber N, Avery A, Takian A, Lichtner V, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from prospective national evaluation in "early adopter" hospitals. BMJ 2011;343:d6054.
[17] Cresswell K, Sheikh A. Electronic health record technology. JAMA 2012;307:2255.
[18] Choi HK, Willett W, Curhan G. Fructose-rich beverages and risk of gout in women. JAMA 2010;304:2270−8.
[19] Williams NM, Zaharieva I, Martin A, Langley K, Mantripragada K, Fossdal R, et al. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. Lancet 2010;376:1401−8.
[20] Gill TM, Allore HG, Gahbauer EA, Murphy TE. Change in disability after hospitalization or restricted activity in older persons. JAMA 2010;304:1919−28.
[21] Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ. What is "quality of evidence" and why is it important to clinicians? BMJ 2008;336:995−8.
[22] Dhingra N, Jha P, Sharma VP, Cohen AA, Jotkar RM, Rodriguez PS, et al. Adult and child malaria mortality in India: a nationally representative mortality survey. Lancet 2010;376:1768−74.
[23] Ludermir AB, Lewis G, Valongueiro SA, de Araujo TV, Araya R. Violence against women by their intimate partner during pregnancy and postnatal depression: a prospective cohort study. Lancet 2010;376:903−10.
[24] Franks PW, Hanson RL, Knowler WC, Sievers ML, Bennett PH, Looker HC. Childhood obesity, other cardiovascular risk factors, and premature death. N Engl J Med 2010;362:485−93.
[25] Prasad V, Cifu A, Ioannidis JP. Reversals of established medical practices: evidence to abandon ship. JAMA 2012;307:37−8.
[26] Prasad V, Cifu A. Medical reversal: why we must raise the bar before adopting new technologies. Yale J Biol Med 2011;84:471−8.
[27] Prasad V, Gall V, Cifu A. The frequency of medical reversal. Arch Intern Med 2011;171:1675−6.
[28] Young NS, Ioannidis JPA, Al-Ubaydli O. Why current publication practices may distort science. PLoS Med 2008;5:e201.
[29] JAMA. JAMA Instructions for Authors. 2012.
[30] Bossuyt PMM. The thin line between hope and hype in biomarker research. JAMA 2011;305:2229−30.
[31] Lumbreras B, Parker LA, Porta M, Pollán M, Ioannidis JPA, Hernández-Aguado I. Overinterpretation of clinical applicability in molecular diagnostic research. Clin Chem 2009;55:786−94.
[32] Greenberg SA. How citation distortions create unfounded authority: analysis of a citation network. BMJ 2009;339:b2680.
[33] Greenberg SA. Understanding belief using citation networks. J Eval Clin Pract 2011;17:389−93.

Appendix 1. Examples of the four types of extrapolations

Article title	What the authors found	The claim made by the authors	Why the medical practice extrapolation may be tenuous
Problems demand solutions	*The authors show that some phenomenon/ topic is a problem and recommend a specific remedy whose benefit is not supported by their data*		
Neuropsychological dysfunction and neuroimaging abnormalities in neurologically intact adults with sickle cell anemia[a]	''Compared with healthy controls, adults with sickle cell anemia (SCA) had poorer cognitive performance, which was associated with anemia and age.''	''First, early identification of patients with difficulties on specific measures of neurocognitive function may allow these patients to enroll in and benefit from cognitive rehabilitation programs.''	Whether cognitive rehabilitation programs will improve cognitive performance in patients with SCA requires prospective study. Whether early identification is necessary to maximize these gains is unknown. How to screen for such patients is a final unanswered question. The current investigation offers no information regarding any of these questions.
Cancer screening among patients with advanced cancer[b]	''A sizeable proportion of patients with (one form of) advanced cancer continue to undergo cancer screening tests that do not have a meaningful likelihood of providing benefit.''	''… as electronic medical records and reminder systems are developed to foster screening adherence, they should also include program features that flag when conditions suggest reevaluation or cessation of screening based on competing comorbidities.''	Although the authors finding reveals a source of wasted health care expenditure, their recommendation for electron medical record warnings is not supported by their investigation and may or may not save health care dollars. Such reminders may be overly burdensome, waste health professionals time, and at a minimum require studies to show that they actually result in fewer wasteful screening tests.
Stillbirth and neonatal death in relation to radiation exposure before conception: a retrospective cohort study[c]	''Uterine and ovarian irradiation significantly increased the risk of stillbirth and neonatal death at doses greater than 1000 Gy [5 (18%) of 28, 9.1 (3.4, 24.6)]. For girls treated before menarche, irradiation of the uterus and ovaries at doses as low as 1.00–2.49 Gy significantly increased the risk of stillbirth or neonatal death [3 (4%) of 69, 4.7 (1.2, 19.0)].''	''Therefore, careful management is warranted for pregnant women treated with high-doses of pelvic irradiation before they have reached puberty.''	The authors have identified a group of patients at risk for a bad outcome; however, they have not shown that ''careful management'' of these patients would make an difference. It would not be surprising if such management (likely referral to maternal fetal medicine) is unable to alter the high-risk outcome, while incurring significant cost.

(Continued)

Appendix 1. Continued

Article title	What the authors found	The claim made by the authors	Why the medical practice extrapolation may be tenuous
Rates of major depressive disorder and clinical outcomes following traumatic brain injury[d]	The authors major finding is that ''among a cohort of patients hospitalized for TBI (traumatic brain injury), 53.1% met criteria for MDD during the first year after TBI.''	''Systematic integration of mental health services into standard care of patients with TBI may be needed to improve long-term outcomes after TBI. Within inpatient rehabilitation, integrated clinical pathways can be used to organize early identification, risk assessment, diagnosis, and guideline-driven treatment of MDD. Systematic depression screening and stepped-care treatment protocols should be integrated into routine outpatient care. For those without or beyond routine follow-up, novel case-finding programs may be useful, possibly via scheduled telephone contacts, Internet-based screening or other technology-assisted methods. The manner in which substance abuse treatment has been integrated into trauma care and depression treatment integrated into primary care may provide models of how to incorporate depression treatment into TBI care.''	Simply because the authors identify a group of patients with a high rate of meeting criteria for MDD, does not mean that surveillance of this group of patients and guidelines based treatment will improve outcomes for these people. A prospective study is needed to show whether or not such strategies benefit patients with TBI. Additionally, screening for depression is a controversial subject, and some groups have eloquently argued why it has not yet shown benefit.[e]
Correlation implies causation	*The authors note that some x is associated with a better (or worse) outcome and conclude that we should (or should not) do (or avoid doing) x*		
Fructose-rich beverages and risk of gout in women[f]	''Compared with consumption of less than 1 serving per month of sugar-sweetened soda, the multivariate relative risk of gout for 1 serving per day was 1.74 [95% confidence interval (CI), 1.19, 2.55] and for 2 or more servings per day was 2.39 (95% CI, 1.34, 4.26) ($P < 0.001$ for trend). The corresponding relative risks for orange juice were 1.41 (95% CI, 1.03, 1.93) and 2.42 (95% CI, 1.27, 4.63) ($P = 0.02$ for trend).''	''Our findings have practical implications for the prevention of gout in women. As conventional dietary recommendations for gout have focused on restriction of purine intake, low-purine diets are often high in carbohydrates, including fructose-rich foods. Our data provide prospective evidence that fructose poses an increased risk of gout among women, thus supporting the importance of reducing fructose intake.''	There are many potential confounders that could explain the authors observation, a lesson shown in hormone replacement therapy and the nurses health study. Simply because fructose intake is associated with a mild increase in gout, does not mean that reducing fructose intake will decrease gout (it may trigger countervailing dietary changes), or that counseling patients about fructose would affect the incidence of gout.
Effect of specialist retrieval teams on outcomes in children admitted to paediatric intensive care units in England and Wales: a retrospective cohort study[g]	The authors studied a large series of children admitted to Pediatric Intensive Care Units (PICUs) in England. 80% of children were retrieved by specialist retrieval teams, and 20% were not. They found that ''In a multivariable analysis, use of a specialist retrieval team for transfer was associated with improved survival (0.58, 0.39–0.87).''	''These findings support the policy of combining centralisation of intensive care services for children with transfer by specialist retrieval teams.''	These findings do support the policy, but there may be other reasons besides the teams themselves that explain the better outcomes obtained by specialist retrieval teams. A prospective randomized study might best adjudicate whether such teams improve outcomes.

Discrimination implies utility	*X can be used to stratify a group of people at risk for some outcome; therefore, we should use x as a s screen*		
Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis[h]	''57 large, rare CNVs (copy number variants) were identified in children with ADHD and 78 in controls, showing a significantly increased rate of CNVs in ADHD (0.156 vs. 0.075; $p = 8.9 \times 10^{-5}$). This increased rate of CNVs was particularly high in those with intellectual disability (0.424; $p = 2.0 \times 10^{-6}$), although there was also a significant excess in cases with no such disability (0.125, $P = 0.0077$).''	''Our results suggest that routine referral to clinical geneticists and screening for such mutations could be helpful for children with ADHD and intellectual disability.''	It is unclear what routine referral and screening would accomplish.
Myocardial fibrosis as an early manifestation of hypertrophic cardiomyopathy[i]	The authors compared biomarker levels in hypertrophic cardiomyopathy patients with particular mutations, with and without LV hypertrophy, and mutation negative normal controls. They found ''Levels of serum C-terminal propeptide of type I procollagen (PICP) were significantly higher in mutation carriers without left ventricular hypertrophy and in subjects with overt hypertrophic cardiomyopathy than in controls (31% and 69% higher, respectively; $P < 0.001$). The ratio of PICP to C-terminal telopeptide of type I collagen was increased only in subjects with overt hypertrophic cardiomyopathy, suggesting that collagen synthesis exceeds degradation.''	''Increased serologic markers of collagen synthesis may identify persons at risk for arrhythmias, sudden death, or heart failure. If so, monitoring levels of these markers may guide new strategies to attenuate disease development or adverse outcomes in hypertrophic cardiomyopathy. We suggest that incorporating genetic testing to identify at-risk mutation carriers, defining features of early disease, and developing therapies to mitigate fibrosis will foster vital new opportunities to change the natural history of hypertrophic cardiomyopathy.''	It remains unclear how to use the information given by genetic testing to guide screening, management or particular therapies. The authors article does not show that genetic testing will benefit this population.
Other assertions	*The authors simply make an assertion on the subject the article addresses*		
Change in disability after hospitalization or restricted activity in older persons[j]	''Hospitalization was strongly associated with … transitions (from health to disability or death), with increased multivariable hazard ratios (HRs) as high as 168 [95% confidence interval (CI), 118, 239] for the transition from no disability to severe disability and decreased HRs as low as 0.41 (95% CI: 0.30, 0.54) for the transition from mild disability to no disability.''	''Given the central role of intervening illnesses and injuries on the disabling process, more aggressive efforts are warranted to prevent their occurrence, to manage them more effectively and reduce subsequent complications, especially in the hospital setting; and, after an event, to enhance restorative interventions in the subacute, home care, and outpatient settings'' and ''more aggressive efforts will be needed to prevent and manage intervening illnesses and injuries, given their apparent role in precipitating and perpetuating the disabling process.''	It is unclear if any of these recommendations will improve the situation where hospitalizations are associated with functional decline or death. Furthermore, the authors' investigation does not make any of these recommendations more or less true or more or less necessary.

(Continued)

Appendix 1. Continued

Article title	What the authors found	The claim made by the authors	Why the medical practice extrapolation may be tenuous
Superficial venous thrombosis (SVT) and venous thromboembolism[k]	''Among 844 patients with SVT at inclusion (median age, 65 yr; 547 women), 210 (24.9%) also had deep venous thrombosis (DVT) or symptomatic pulmonary embolism.''(PE)	''Our findings also suggest that compression ultrasonography might be considered for patients with symptomatic SVT at presentation to evaluate the extent of the thrombosis and diagnose potential DVT, that physicians should suspect and test for pulmonary embolism in patients with suggestive symptoms, and that close follow-up of patients with isolated SVT might be advisable to detect early complications that involve the deep veins.''	The authors' article does not show whether screening these patients with compression ultrasonography will be beneficial. This is a hypothesis best tested by a future prospective study. Their recommendation that physicians should suspect and test for PE in those with symptoms suggestive of PE offers little novel advice.

[a] Vichinsky EP, Neumayr LD, Gold JI, Weiner MW, Rule RR, Truran D, et al. Neuropsychological dysfunction and neuroimaging abnormalities in neurologically intact adults with sickle cell anemia. JAMA 2010;303:1823–1831.

[b] Sima CS, Panageas KS, Schrag D. Cancer screening among patients with advanced cancer. JAMA 2010;304(14):1584–1591.

[c] Signorello LB, Mulvihill JJ, Green DM, Munro HM, Stovall M, Weathers RE, et al. Stillbirth and neonatal death in relation to radiation exposure before conception: a retrospective cohort study. Lancet 2010;376:624–630.

[d] Bombardier CH, Fann JR, Temkin NR, Esselman PC, Barber J, Dikmen SS. Rates of major depressive disorder and clinical outcomes following traumatic brain injury. JAMA 2010;303:1938–1945.

[e] Thombs BD, Coyne JC, Cuijpers P, De Jonge P, Gilbody S, Ioannidis JPA, et al. Rethinking recommendations for screening for depression in primary care. CMAJ 2011; DOI:10:1503/cmaj.111035. 2.

[f] Choi HK, Willett W, Curhan G. Fructose-rich beverages and risk of gout in women. JAMA 2010;304(20):2270–2278.

[g] Ramnarayan P, Thiru K, Parslow RC, Harrison DA, Draper ES, Rowan KM. Effect of specialist retrieval teams on outcomes in children admitted to paediatric intensive care units in England and Wales: a retrospective cohort study. Lancet 2010;376:698–704.

[h] Williams NM, Zaharieva I, Martin A, Langley K, Mantripragada K, Fossdal R, et al. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. Lancet 2010;376(9750):1401–1408.

[i] Ho CY, López B, Coelho-Filho OR, Lakdawala NK, Cirino AI, Jarolim P, et al. 2010. Myocardial fibrosis as an early manifestation of hypertrophic cardiomyopathy. N. Engl J. Med. 363;552–563.

[j] Gill TM, Allore HG, Gahbauer EA, Murphy TE. Change in disability after hospitalization or restricted activity in older persons. JAMA 2010;304(17):1919–1928.

[k] Decousus H, Quere I, Presles E, Becker F, Barrellier MT, Chanut M, et al. Superficial venous thrombosis and venous thromboembolism: a large, prospective epidemiologic study. Ann Intern Med 2010;152:218–224.