PERSPECTIVES

Could Media Reports Include a Standardized Scale for Quality of Evidence?

R. Scott Braithwaite, MD, MS, FACP

Section of Value and Comparative Effectiveness, Division of General Internal Medicine, New York University School of Medicine, New York, NY, USA.

While some media reports offer accurate interpretations of clinical research, other reports are misleading. The uneven accuracy of medical reporting may act in concert with its sheer volume to confuse the lay public about which health messages are most important and evidence-based. I outline one possible step towards a solution: medical journals can embed quality of evidence ratings in article summaries and create incentives for inclusion of these ratings in lay media reports.

KEY WORDS: health care journalism; medical journalism; media; news; quality of evidence.

J Gen Intern Med 26(5):543–5

DOI: 10.1007/s11606-010-1599-z

© Society of General Internal Medicine 2010

While some media reports offer accurate and circumspect lay translations of clinical research, other reports are misleading. Some headlines have included “Want to Cut Cancer Risk? Try Munching Pistachios,” “New migraine cure? Forehead lift surgery helps patients,” and “Urine Test May Spot Sleep Apnea in Children.”¹ According to journalist Susan Dentzer “Journalists sometimes feel the need to play carnival barkers, hyping a story to draw attention to it. This leads them to frame a story as new or different—depicting study results as counterintuitive or a break from the past—if they want it to be featured prominently or even accepted by an editor at all.”² Published reports suggest that media descriptions of clinical research omit most relevant items and have shown little or no quality improvement over time.^{3–6}

The uneven accuracy of medical reporting may act in concert with its sheer volume to desensitize an information-weary public. Like clinicians who tune out a life-saving electronic clinical reminder because they are inundated by false alarms, an individual who hears too many stories about lung cancer risk (e.g., that pistachios increase lung cancer risk) may respond by tuning out the entire barrage of cancer-related information. As a result, she may not pay attention to information that could substantially improve her health and is backed by strong evidence. In the current US health environment, where much preventable morbidity is caused by smoking, and excessive alcohol and excessive caloric intake, sensationalistic reporting may mute important messages and thereby harm public health. When everything is important, nothing is important.

A common reason why much media reportage is misleading is an inappropriate inference of causality. Indeed, some “watchdog” groups such as www.healthnewsreview.org survey how causality is portrayed in the media;¹ and the Association of Health Care Journalists advises journalists to “avoid vague, sensational language” and to “understand the process of medical research in order to report it accurately.”⁷ Demonstration of an association in one small study carries very different implications than data from several large randomized controlled trials, yet similar headlines may be used, leaving the public without a frame of reference.

DEVELOPING AN EVIDENCE SCALE FOR A LAY AUDIENCE

Decision makers in health policy and clinical care often need to synthesize studies with greatly varying strength of evidence to clarify the scope of inference and to evaluate whether it is a sound basis for decision making. Indeed, scientific reports often employ standardized evidence scales, such as that used by the United States Preventive Services Task Force (USPSTF).^{8,9} While the relative merits of individual scales are widely debated, evidence scales have face validity, and are becoming increasingly ubiquitous in the reports of expert panels and clinical guidelines.

However, despite the promulgation of evidence scales for clinicians and other health professionals, there is no corresponding evidence scale developed with a lay audience in mind. Because many individuals are familiar with letter and numerical “grades,” it seems likely that a lay audience could interpret a transparent, unidimensional scale that is intuitively rank-ordered (say, “A” through “D”). Indeed, the popularity of distilling complicated quality of care metrics into standard rank-order grading systems (for example, the National Committee for Quality Assurance assigns numbers of stars on a scale from 0 to 4)¹⁰ provides support of this idea.

While it is debatable which evidence scale would be best suited for media reports, the USPSTF rating scale offers one possible starting point (Table 1). It could be adapted for communicating strength of evidence to a lay audience by mapping each letter grade to a corresponding concise, standardized lay explanation. This modified USPSTF scale could then be embedded within media reports that describe new studies, and could communicate (1) the level of evidence prior to the study, (2) how the study changed the level of evidence (if at all), and (3) the level of evidence after the study’s results are considered together with prior studies (Table 2). A summary evidence table in a recent New York

Received August 24, 2010

Accepted November 19, 2010

Published online December 21, 2010

Table 1. Potential Adaptation for USPSTF Evidence Scale Along with Equivalent Lay Explanations

USPSTF grade	USPSTF definition	Clinical explanation	Lay explanation
A	The USPSTF recommends the service. There is high certainty that the net benefit is substantial	Offer or provide this service	PROVEN HEALTH EFFECT: Scientific evidence suggests important link with high certainty
B	The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial	Offer or provide this service	POSSIBLE HEALTH EFFECT: Scientific evidence suggests link
C	The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small	Offer or provide this service only if other considerations support offering or providing the service to an individual patient	HEALTH EFFECT VARIES FROM PERSON TO PERSON: Scientific evidence suggests that personalized interpretation is necessary because there is a "close call" of benefits and harms, or because benefits and/or harms vary greatly from person to person
D	The USPSTF recommends not using the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits	Discourage the use of this service	PROVEN TO HAVE NO HEALTH EFFECT: Scientific evidence suggests there is no link
I Statement	The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined	Read the clinical considerations section of USPSTF Recommendation Statement. If the service is offered, patients should understand the uncertainty about the balance of benefits and harms	NO PROVEN HEALTH EFFECT: Scientific evidence is inconclusive or insufficient

Times article on Alzheimer's disease¹¹ included many of these attributes and offers support for the feasibility of this approach. Growing enthusiasm surrounding the "Drug Facts Box" (e.g., a transparent and decision-centered lay summary of a drug's benefits and harms) also highlights the virtues of providing lay evidence summaries in standardized and decision-ready formats.¹²

ENCOURAGING USE OF AN EVIDENCE SCALE FOR A LAY AUDIENCE

Embedding standardized evidence grades in media reporting is unlikely to happen spontaneously. More transparent evidence reporting may place the limitations of scientific articles in sharper focus, rendering them less newsworthy.¹³ The first journal that adapts such a practice may find less interest in its releases, and the first media organization that adapts such a practice may find less interest in its reports.

The primary purpose of this article is to stimulate debate about the appropriateness of incorporating evidence scales in lay media reports rather than to outline a particular path forward. Nonetheless, I would like to outline one approach that may be plausible. A coalition of medical journals could propose to provide earlier press releases to those media outlets that regularly incorporate lay evidence scales. Their proposal would need to be attractive to high-profile media outlets, offering a sufficient "reward" to confer a strong competitive advantage. The coalition of journals advancing this proposal would need to publish enough newsworthy articles so that this competitive advantage would be especially compelling. If this proposal was successfully adopted, it might then spread spontaneously beyond a small group of early adopters, as more media outlets would seek early access to the most newsworthy medical research, and other journals would become convinced of the feasibility of this approach. Gradually, as the use of standardized evidence scales became more routine, the lay audience might become accustomed to hearing health information in a new decision-relevant context.

Table 2. Example of How Use of a Modified Evidence Scale Could Summarize a Single-Site, Small, Intervention Study Finding that Oral Pistachio Intake Raises Serum Levels of Gamma-Tocopherol (Which Had Been Associated with Lower Risks of Several Non-Lung-Cancers in Prior Cross-Sectional Studies)

Evidence of link between pistachios and lung cancer	Before study was done	What study showed	After study was done
	I	I	I

A=PROVEN HEALTH EFFECT: Scientific evidence suggests important link with high certainty

B=POSSIBLE HEALTH EFFECT: Scientific evidence suggests link

C=HEALTH EFFECT VARIES FROM PERSON TO PERSON: Scientific evidence suggests that personalized interpretation is necessary because there is a "close call" of benefits and harms, or because benefits and/or harms vary greatly from person to person

D=PROVEN TO HAVE NO HEALTH EFFECT: Scientific evidence suggests there is no link

I=NO PROVEN HEALTH EFFECT: Scientific evidence is inconclusive or insufficient

One might ask why any medical journals would ever propose to take on the additional responsibility of summarizing evidence while running the risk that many of their reports would get less rather than more media coverage. Indeed, press releases from medical journals often ignore limitations in order to garner more media coverage.¹⁴ However, it is important to note that medical journals have multiple priorities, and these priorities include truthful dissemination of their scientific reports and promotion of public health. Leading medical journals that already demonstrate their commitment to these priorities by providing general summaries of the scientific articles that they publish (e.g., "Summary for Patients" in *Annals of Internal Medicine* or "Editors' Summary" in *PLoS Medicine*) may be willing to lead a coalition to improve lay evidence reporting. In addition, advantages may accrue to early adopters. Journals may be more certain of getting high-profile media coverage of the top tier evidence that they publish, and may gain prestige through early adoption of a standard that leads to more truthful reporting and dissemination of medical research.

PITFALLS OF INCORPORATING EVIDENCE SCALES INTO MEDIA REPORTS

Some may argue that a single letter grade is insufficient for evaluating evidence, because its very uni-dimensionality fails to capture the array of important attributes necessary for evidence interpretation, such as the magnitude of effect and statistical uncertainty. However, an analogous argument may be applied to virtually any other source of lay information that employs single letter or number grades, yet these grades are often embraced by their intended audience. For example, the quality of health care organizations clearly has distinct domains, yet the National Committee for Quality Assurance chooses to synthesize individual, domain-specific grades into a summary grade. Consumer products have individual attributes, yet the Consumers Union chooses to synthesize individual, attribute-specific ratings into a single rating. It is important to observe that providing a summary grade does not preclude the option of supplementing it with grades for individual evidence attributes.

It also may be argued that editors write headlines rather than journalists, and so incorporation of an evidence scale is unlikely to reduce the number of sensationalist headlines. While this argument may be true in the short term, it is less likely to be true in the long term, as an increasing lay understanding of evidence ratings could make a sensationalistic headline more incompatible with the article that follows.

Because the "gray literature" (e.g., blogs) is becoming a more important source of lay medical information, it would be desirable to include these sources in any systematic effort to make evidence more transparent. Yet the approach advocated here may be less feasible for gray literature sources than for established scientific journals.

Editorial staff of medical journals are often overwhelmed by existing work demands and may balk at assuming the additional responsibility of not only evaluating the merit of the research, but also summarizing its addition to the cumulative body of evidence. However, over time, the authors and peer-reviewers may absorb some of this additional burden, especially if editors require an evidence summary to be incorporated into the Discussion section of the manuscript. The work burden

could also be restricted by limiting lay evidence reports to those articles that are particularly relevant to public health (e.g., when high prevalence conditions or risk factors are being associated with diseases that confer great morbidity and mortality).

Finally, it could be argued that evidence ratings should always be performed by impartial organizations that have nothing to gain by grade inflation (e.g., the USPSTF itself). However, it seems implausible that any organization would ever be sufficiently resourced or invested with sufficient authority to offer prompt evaluation of all research reports at the time of press.

CONCLUSION

While there are many challenges to incorporating strength of evidence ratings into media reports, efforts to improve the signal-to-noise ratio of health information may ultimately motivate more targeted and beneficial changes in health behaviors, and lead to improved public health.

Conflict of Interest: None disclosed.

Corresponding Author: R. Scott Braithwaite, MD, MS, FACP, Section of Value and Comparative Effectiveness, Division of General Internal Medicine, New York University School of Medicine, 423 East 23rd Street, 15th Floor, New York, NY 10010, USA (e-mail: scott.braithwaite@nyumc.org).

REFERENCES

1. **Health News Review**—Holding health and medical journalism accountable. Available at: www.healthnewsreview.org. Accessed November 22, 2010.
2. **Dentzer S.** Communicating medical news—pitfalls of health care journalism. *N Engl J Med.* 2009;360:1–3.
3. **Wilson A, Bonevski B, Jones A, Henry D.** Media reporting of health interventions: signs of improvement, but major problems persist. *PLoS ONE.* 2009;4:e4831.
4. **Bonevski B, Wilson A, Henry DA.** An analysis of news media coverage of complementary and alternative medicine. *PLoS ONE.* 2008;3:e2406.
5. **Holtzman NA, Bernhardt BA, Mountcastle-Shah E, Rodgers JE, Tambor E, Geller G.** The quality of media reports on discoveries related to human genetic diseases. *Community Genet.* 2005;8:133–44.
6. **MacKenzie R, Chapman S, Barratt A, Holding S.** "The news is [not] all good": misrepresentations and inaccuracies in Australian news media reports on prostate cancer screening. *Med J Aust.* 2007;187:507–10.
7. **Schwitzer G.** A statement of principles for health care journalists. *Am J Bioeth.* 2004;4:W9–13.
8. **Barton MB, Miller T, Wolff T, et al.** How to read the new recommendation statement: methods update from the US Preventive Services Task Force. *Ann Intern Med.* 2007;147:123–7.
9. **Braithwaite RS, Roberts MS, Justice AC.** Incorporating quality of evidence into decision analytic modeling. *Ann Intern Med.* 2007;146:133–41.
10. **National Committee for Quality Assurance.** Available at: <http://reportcard.ncqa.org/plan/external/Plansearch.aspx>. Accessed November 22, 2010.
11. **Kolata G.** Years later, no magic bullet against Alzheimer's Disease. *New York Times* August 29, 2010.
12. **Schwartz LM, Woloshin S, Welch HG.** Using a drug facts box to communicate drug benefits and harms: two randomized trials. *Ann Intern Med.* 2009;150:516–27.
13. **Anonymous.** Does the media support or sabotage health? *Lancet* 2009;373:604.
14. **Woloshin S, Schwartz LM.** Press releases: translating research into news. *JAMA.* 2002;287:2856–8.