When patient-centeredness and evidence-based medicine collide

R Scott Braithwaite

Division of Comparative Effectiveness and Decision Science, New York University School of Medicine, New York, NY, USA

Abstract: There are numerous clinical situations in which inferences from evidence-based medicine conflict with patient-reported outcomes or experiences. For example, a patient may report better symptom relief from a drug which has been demonstrated in randomized controlled trials to be non-superior to its competitors. Such conflicts have often been cast as tensions between "the art of medicine" and the "the science of medicine". However, we add to the current evidence-based literature by asking whether many current distinctions between "the art of medicine" versus the "science of medicine" are not best explicated in those terms after all, but rather are proxy terms for whether internal validity or external validity are more important in a particular situation. In addition, we outline one possible framework for systematically determining whether evidence is generalizable in a particular clinical situation. Limitations to this approach are emphasized, as well as steps forward that would make use of published but underutilized methods.

Keywords: art of medicine, science of medicine, internal validity, external validity, patient-reported, generalizability

Introduction

The goal of science is creating "generalizable" knowledge; the goal of medicine is best treating individual patients. However, "generalizable" inferences from evidence-based medicine (EBM) may conflict with patient-reported outcomes or experiences. For example, consider a hypothetical case: Georgine Banks, a 54-year-old female with osteoarthritis, gastroesophageal reflux disease, diabetes, and obesity, has just joined your patient panel. Her previous practitioner had prescribed lamotrigine for her neuropathic pain, which was her most function-limiting symptom and interfered with her work and many of her hobbies. She reports that lamotrigine improves her neuropathy symptoms dramatically without side effects, allowing her to resume normal activities. Even after you acquaint her with the scientific evidence behind the superiority of amitriptyline and duloxetine to lamotrigine, she remains unimpressed, and wants to continue with her lamotrigine.

The art of medicine versus the science of medicine: a false dichotomy?

Even the most committed aficionados of EBM may find themselves leaving Mrs Banks on lamotrigine. But then an EBM practitioner would be flouting a clear application of EBM. Why? While similar conflicts have been cast as a tension between the "art" and "science" of medicine, I ask whether they could be illuminated within a scientific

Correspondence: R Scott Braithwaite
Division of Comparative Effectiveness and Decision Sciences, Department of Population Health, NYU School of Medicine, 227 East 30th street, Floor 6 Room 615, New York, NY 10010, USA
Tel +1 212 263 4964
Email scott.braithwaite@nyumc.org

context as a tension between the concept of "internal validity" (that is, more likely to represent truth in the sample of patients studied) and the concept of "generalizability" (that is, more likely to represent truth outside the experimental context). While generalizability is often felt to be optimized by the pragmatic approach of designing trials with broad inclusion criteria such that results should "generalize" to most patients seen in routine care, here, we are referring to a specific aspect of generalizability perhaps better described as "applicability".[1] "Applicability" emphasizes the fitness of the data to the specific target of inference in which a particular evidence-based decision needs to be made – in this case the specific decision of the best therapy for Mrs Banks. The tension between internal validity and applicability can be suppressed if an evidence synthesis subordinates applicability to internal validity, as is common in evidence synthesis methods that are commonly employed (for example, GRADE [Grading of Recommendations Assessment, Development,

and Evaluation]). However, it looms over treatment decisions like that for Mrs Banks. Table 1 offers one of many possible ways of reconciling evidence regarding internal validity with evidence regarding applicability, explicitly considering heterogeneity of treatment effect. This table is not meant to suggest that the illustrated approach is unique or superior to others, but rather to illustrate that a far more systematic approach to evidence application is possible.

Internal validity versus applicability

Assume Mrs Banks' clinician performs a PICO (Patient, Intervention, Control, Outcome) evidence synthesis of lamotrigine versus amitriptyline: the patient is Georgine Banks and her primary presenting problem is pain in her extremities due to diabetic neuropathy; the intervention is lamotrigine (varying doses); the control is amitriptyline (varying doses); and the outcomes are neuropathic pain control and functional limitation. Consider two alternative evidence syntheses, one

Table 1 A hypothetical, systematic approach for reconciling the "art of medicine" (eg, strength of applicability, a component of external validity) and the "science of medicine" (eg, strength of internal validity), and recognizing the importance of Bayesian priors in interpreting subgroup analysis

Scenario	Patient characteristic(s) reflected in subgroup analyses?	Patient characteristic(s) a priori may result in clinically significant possibility that harm may exceed benefits (eg, Bayesian prior)	Applicability to patient decision	Possible evidence grade
Patient has no characteristics that were exclusion criteria in trial	Unfavorable HTE detected	Yes	No	D
		No	No	D
	HTE not detected or favorable HTE detected	Yes	Consider with extreme caution	C
		No	Yes	A
Patient has ≥1 characteristics that were exclusion criteria in trial	N/A	Yes	Consider with extreme caution	C
		No	Consider	B
Patient has ≥1 characteristics that were subject to subgroup analyses in trial, but without sufficient statistical power to detect heterogeneity of treatment effect (ie, exploratory analyses)[8]	Unfavorable HTE detected	Yes	No	D
		No	No	D
	HTE not detected or favorable HTE detected	Yes	Consider	B
		No	Yes	A
Patient has ≥1 characteristics that were subject to subgroup analysis in trial, but with sufficient statistical power to detect heterogeneity of treatment effect (ie, confirmatory analyses)[8]	Unfavorable HTE detected	Yes	No	D
		No	No	D
	HTE not detected or favorable HTE detected	Yes	Yes	A
		No	Yes	A
Patient has direct experience contradicting scientific evidence, anecdotally positive	N/A	Yes	Consider with extreme caution	C
Patient has direct experience contradicting scientific evidence, anecdotally negative	N/A	No	No	D
Patient has direct experience contradicting scientific evidence, scientifically positive	N/A	Yes	Consider	B
Patient has direct experience contradicting scientific evidence, scientifically negative	N/A	No	No	D

Note: This table is not meant to suggest that the illustrated approach is unique or superior to others, but rather to illustrate that a far more systematic approach to evidence application is possible.
Abbreviations: HTE, heterogeneity of treatment effect; N/A, not applicable.

prioritizing "applicability" and one prioritizing "internal validity". The synthesis prioritizing applicability emphasizes data from the particular patient to which the decision applies. Conversely, the synthesis prioritizing internal validity emphasizes studies meeting prescribed methodological standards for internal validity (eg, GRADE), even if they include (as inevitably they must) patients dissimilar to Mrs Banks who may have a dissimilar treatment response. The synthesis emphasizing internal validity would suggest that Mrs Banks should not receive lamotrigine, relying on evidence with high internal validity (an experimental, blinded, prospective assessment of a validated outcome measure). Conversely, the synthesis emphasizing applicability would suggest that Mrs Banks should receive lamotrigine. Mrs Banks' pain relief is of the highest possible applicability because it applies directly to her, even though the underlying data has abhorrent internal validity (non-experimental, non-blinded, retrospective assessment of a subjective, non-validated outcome measure). Consequently, alternative evidence syntheses lead to opposite inferences for treating Mrs Banks.

Heterogeneity of treatment effect is the underlying reason why inferences based on emphasizing internal validity may conflict with inferences based on emphasizing applicability and can represent a challenge to the standardization of practice, especially when the predictors of response heterogeneity are unknown. While some have questioned the existence of individualized responses,[2] cross-over studies have shown that half of patients unresponsive to a particular antidepressant may respond to another drug in the same class.[3–5] This challenge should be accepted rather than wished away via inflexible formulary designs. Drugs showing equal efficacy on average may have considerable variability in individual patients, and may therefore not always be therapeutically interchangeable. Further, treatments that are inferior on average, may be better for some (like Mrs Banks).

Could an emerging era of increased HIT sophistication lead to higher quality data on applicability?

Internal validity typically trumps applicability in evidence hierarchies when the two conflict because of concerns about biased data. But what if decisions could be based on highly applicable data that were not biased? Comparisons of patient experience under different therapies have always been a cornerstone of clinical care. Emerging health information technology (HIT) may nudge clinicians to reimagine state-of-the-art clinical observation as including rigorous and quantitative attention to patient-reported outcomes. Indeed, data collected in routine clinical care and stored in HIT could be harnessed to conduct patient-based evidence reviews, and could be integrated with sample-derived data to inform decision making. While evidence reviews that maximize applicability may produce low-quality evidence, scientific methods exist to improve their quality, even though these methods are seldom employed in practice. For example, Mrs Banks could be the subject of a cross-over N=1 trial with double-blinding, placebo control, use of validated measurement scales for symptom control, and duration and cross-over frequency based on formal tests of statistical significance. Indeed, N=1 trials could hypothetically be conducted by well-trained independent laboratories as if they were diagnostic tests.

That being said, there will always be important limitations to collecting higher quality patient-based data. Highly valid patient-centered evidence (eg, N=1 trials) would often be infeasible, particularly if no psychometrically validated outcome measures exist, outcomes are infrequent, long time delays occur before benefits or harms manifest, blinding is impractical, or a patient is unwilling to enroll in a trial or to take a placebo. In addition, they may be invalid if a condition is not stable, if long time delays occur before benefits or harms manifest, or if placebo effects overwhelm biological effects. Because resources are limited, gathering evidence of this rigor would be unsuitable unless the value of information exceeded its costs.[6]

While few might deny that a highly rigorous, blinded, randomized N=1 trial for Mrs Banks would provide stronger evidence than a conventional RCT in others for treatment decisions in Mrs Banks herself, what if some of the study standards were relaxed? What if, for example, Mrs Banks was unblinded? Wouldn't that provide a "pragmatic" result for Mrs Banks, applicable to the real world in which she knows what treatment she receives? And if the trial was unblinded, how critical is randomization, since comparability in an N=1 trial (where Mrs Banks acts as her own control) is less an issue than in conventional trials? At what point, then, does careful clinical observation cease to be science?

Limitations

It may be argued by some that no tension exists between evidence-based medicine and patient-centeredness as long as a practitioner adheres to a sufficiently broad interpretation of evidence-based medicine, such as the famous definition of Sackett et al as "the conscientious, explicit and judicious use of current best evidence in making decisions about the

care of individual patients".[7] However, an excessively broad interpretation of evidence-based medicine runs the risk of being too vague to guide clinical decision making.

Some contend that there is nothing fundamentally antagonistic to EBM about keeping a patient on a medicine that is working for them when the risks are outweighed by the benefits, regardless of the evidence in randomized controlled trials. However, this argument may not be valid when evidence suggests that an alternative therapy may offer greater benefits and/or lesser risks. Also, preserving patient autonomy may sometimes "trump" the application of any evidence hierarchy from an ethical perspective, as long as harms do not exceed benefits, particularly if a patient is more activated and/or has a greater propensity to adhere when she influences the clinical decision.

Finally, it may be argued that the term "art of medicine" is obsolete and therefore the discussion underlying this article is irrelevant. However, I hear "the art of medicine" used daily in the context of clinical discussions across multiple settings, so my anecdotal experience (albeit of low evidentiary quality) advises against this argument.

Conclusion

As alternative evidence syntheses may reach opposing conclusions, it remains controversial which medication will best control Mrs Banks' future symptoms – the one that she reports helped her in the past, or the one that seems to work better on average in other people. It should be pointed out, however, that while this remains an unanswered question, it is not scientifically unaddressable. When should patient-specific evidence be gathered systematically to supplement sample-based data? Should we apply sample-based EBM when conflicting, maximally applicable patient-centered data emerges? These questions should be debated explicitly and explored scientifically, rather than cast as a battle of perspectives between "medicine as science" and "medicine as art".

References

1. Atkins D, Chang SM, Gartlehner G, et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. *J Clin Epidemiol*. 2011;64(11):1198–1207.
2. Senn S. Individual response to treatment: is it a valid assumption? *BMJ*. 2004;329(7472):966–968.
3. Zarate CA, Kando JC, Tohen M, Weiss MK, Cole JO. Does intolerance or lack of response with fluoxetine predict the same will happen with sertraline? *J Clin Psychiatry*. 1996;57(2):67–71.
4. Thase ME, Blomgren SL, Birkett MA, Apter JT, Tepner RG. Fluoxetine treatment of patients with major depressive disorder who failed initial treatment with sertraline. *J Clin Psychiatry*. 1997;58(1):16–21.
5. Simon G. Choosing a first-line antidepressant: equal on average does not mean equal for everyone. *JAMA*. 2001;286(23):3003–3004.
6. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ*. 1996;5(6):513–524.
7. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996; 312:71.
8. Varadhan R, Stuart EA, Louis TA, Segal JB, Weiss CO. *Review of Guidance Documents for Selected Methods in Patient Centered Outcomes Research: Standards in Addressing Heterogeneity of Treatment Effectiveness in Observational and Experimental Patient Centered Outcomes Research. A Report to the PCORI Methodology Committee Research Methods Working Group*. Washington, DC: Patient-Centered Outcomes Research Institute; 2012. Avaialble from: http://www.pcori. org/assets/Standards-in-Addressing-Heterogeneity-of-Treatment-Effectiveness-in-Observational-and-Experimental-Patient-Centered-Outcomes-Research.pdf. Accessed May 23, 2014.