ELSEVIER

Journal of Clinical Epidemiology 61 (2008) 521−524

Journal of Clinical Epidemiology

Value judgments in the analysis and synthesis of evidence

Daniel Strech[a,*], Jon Tilburt[b]

[a]Institute of Medical Ethics, University of Tübingen, Germany
[b]Division of General Internal Medicine and the Bioethics Research Program, Mayo Clinic, Rochester, MN, USA

Accepted 2 January 2008

Abstract

Objective: To describe the principal role of value judgments in the analysis and synthesis of evidence as they are involved in systematic reviews, meta-analyses, and health technology assessments.

Method: Using the tools of conceptual analysis, we characterize three main types of value judgments and propose an outline of how to enhance the appropriate role of value judgments in the process of analyzing and synthesizing evidence.

Results: The production, analysis, and synthesis of evidence involve value judgments characterized as preferences of persons or groups that cannot be validated by appeal to facts alone. Because preferences across individuals can vary, value judgments can be a source of bias in science and unwarranted variation in the application of scientific evidence. However, it is not possible or desirable to eliminate all value judgments in the process from production to synthesis of evidence.

Conclusion: With respect to the value judgments that shape the analysis and synthesis of evidence, review authors should disclose and justify choices related to the three key value judgments outlined in this paper. Authors should also highlight how their value judgments differ from the stated or implicit value judgments of previously published reviews on the same topic.

Keywords: Bioethics; Value judgments; Evidence-based medicine; Bias; Transparency; Systemic reviews

1. Evidence and value judgments

The production, analysis, and synthesis of evidence are scientific endeavors. Science seeks a better understanding of truth. In clinical biomedical science, this quest results in attempts to make accurate estimates of risk and benefit of health interventions to better approximate truths about health. In seeking the truth, scientists must take measures to minimize bias. The tools of biomedical research and evidence-based medicine help both designers and users of the evidence to minimize exposure to biases [1−4]. There also are well-described biases in the interpretation and communication of evidence [5,6].

Scientific evidence supports the search for accurate descriptions of what *is* the case. However, for science to influence decision making, descriptive ''is'' statements be combined with value judgments (VJs). VJs are preferences of persons or groups that cannot be validated by appeal to facts alone. They are a separate category of human reasoning that complements data from observation. Because preferences across individuals can vary, VJs can be a source of bias in science and unwarranted variation in the application of scientific evidence. However, all VJs in science do not necessarily result in bias. And although producers and interpreters of biomedical science should strive to eliminate bias, it is simply impossible and undesirable to eliminate all VJs in the production, analysis, synthesis, and application of evidence. This has already been recognized in different steps of the *production* of evidence such as determining adequate alpha errors or sample sizes, assessing content validity for quality-of-life scales, or incorporating patient preferences into randomized trials [7−10]. Other authors have called attention to the importance of VJs in the *application* of evidence. For instance, determining the levels of significance or thresholds for cost-effectiveness requires VJs [10−12]. However, there is still little awareness of VJs in the analysis and synthesis of evidence such as occurs in systematic reviews. In a recent interview with Sean Tunis, David Eddy argues that scientific judgments and VJs are rather separate [13]. He states that in the analysis of evidence we use scientific judgments, whereas later on, when applying the evidence to individuals or health policy decisions, VJs come into play. In contrast, here we contend that along the complete spectrum from producing, analyzing, synthesizing, and applying evidence, investigators face

* Corresponding author. Institute of Medical Ethics, University of Tübingen, Schleichstraße 8, 72076 Tübingen, Germany. Tel.: +49-7071-29-78016; fax: +49-7071-29-5190.
E-mail address: daniel.strech@uni-tuebingen.de (D. Strech).

0895-4356/08/$ − see front matter
doi: 10.1016/j.jclinepi.2008.01.001

VJs. If so, those VJs deserve explicit examination. Unlike earlier work we focus more specifically on VJs that occur in the process of evidence analysis and synthesis as occurs commonly in systematic reviews, meta-analyses, and health technology assessments. In so doing, we hope to further promote the dialogue about the role of VJs in clinical medical sciences.

As we outline in Table 1, VJs come into play in most steps along the spectrum of evidence production, analysis, synthesis, and application. The GRADE approach and the CONSORT statement give a more detailed explanation of these steps [3,14]. Whenever scientists use words like ''appropriate,'' ''sufficient,'' ''significant'' in presenting or appraising evidence, they are making VJs. Which level of significance is *adequate*? Which criteria are *appropriate* to judge the study quality? Are the inconsistencies within the study or among studies *important*? *Should* the study be excluded from meta-analysis? Despite their importance, the role of VJs particularly in the analysis and synthesis of evidence has not been discussed in detail.

Below we propose a categorization for three main VJs in the analysis and synthesis of evidence. Table 2 presents three basic types of VJs (Table 2). These general categories may not capture all VJs that come into play in the analysis and synthesis of evidence, but they do begin to clarify an important and underappreciated dimension of these processes.

2. The framework of value judgments in the analysis and synthesis of evidence

We can distinguish between three principal types of VJs in the analysis and synthesis of evidence: judgments about (1) choosing outcome measures, (2) balancing benefits and harms, and (3) tolerating uncertainty.

2.1. Choosing outcome measures

Clinical research is only as good as its outcome measures. In order for research to be scientifically valid investigators must specify in advance the outcomes of interest for their study. Because the effects of interventions on diseases can be described and characterized by different outcome measures, choosing the best and most feasible outcome measure is essential to answering the research question. Increasingly, patient reported measures, such as pain, sadness, or quality of life, have taken on greater importance as outcome measures [15]. Similarly, investigators must determine if surrogate outcome measures are warranted. In diabetes research, for instance, investigators must decide if hemoglobin A1C can or should be substituted for more definitive outcome measures such as progression to blindness, renal failure or death. In general, it is also important to minimize the number of outcomes measured in a study as well as in a systematic review for analytic reasons.

VJs inevitably contribute to decisions of which outcome measures to include in a systematic review, meta-analysis, or health technology assessment. When review authors decide which outcomes are more or less important they are making VJs. Should mortality be the primary outcome? Or should quality of life? Are surrogate outcome measures acceptable for the specific research context? Answering these questions requires sound clinical and statistical knowledge as well as a final choice about the outcome to be used. Although the preference for outcome measures is usually not an arbitrary one, it nevertheless represents a preference that others may disagree with and which requires some further justification. Often the justification for choosing an outcome in the original study design boils down to a process of balancing scientific and practical considerations, such as resources, effort, and time. This balancing process depends on VJs reflecting the importance the

Table 1
Judgments in the production of evidence and their normative characteristics

Judgments in the production, analysis, and synthesis of evidence	Examples (normative characteristics)
Conceptualization and simplification of patient relevant outcomes	What are the *relevant* outcomes?
	What degree of simplification in outcome measurement is warranted?
Prioritization of research question	Which research question is *worth* studying and *should* be financed?
Study design	Which study design is *appropriate* for the specific clinical question?
Importance of outcomes	Which outcomes (for benefits and harms) are more and which are less *important*?
Analysis	Which level of significance is *adequate* to the research question?
Reporting	Which degree of simplification within the reporting of study findings is *appropriate*? Which findings *should* be explicitly reported and which can be ignored? Which models of reporting the effect size are *appropriate*?
Study quality (selection and specification of assessment criteria and consistency in data)	Which criteria are *appropriate* to judge the study quality?
	Are the inconsistencies within the study or in relation to other studies *important*?
	Which limitations/shortcomings in the study quality are *sufficient for excluding or discounting the results*?
Data synthesis (inclusion and exclusion of data)	Is the method for data aggregation *appropriate*?
	Should the study be excluded of meta-analysis?
Balancing benefits and harms	Are the benefits of the intervention are *more important* than the harms and vice versa?
Strength of recommendations (overall quality, downgrading)	What are the *adequate* reference points (outcomes for benefits and harms) to judge the overall quality? What framework for the ''level of evidence'' and ''strength of recommendation'' is *appropriate*?

Table 2
Value judgments and evidence synthesis

Three key value judgments
Choosing outcome measures—Which outcome measures are more or less important for the patient and for the evaluation of a certain intervention?
Balancing benefits and harms—Given the magnitude of effect for a given outcome and the quality of those research results, does the intervention do more good than harm?
Tolerating uncertainty—Deciding the merits of gaining more but uncertain knowledge vs. gaining less knowledge with greater certainty

investigator places on each scientific and practical factor under consideration. Evaluating the merits of different outcome measures occurs at several points along the spectrum from evidence production to evidence synthesis. For instance, when designing a clinical trial, an analysis plan specifies in advance what the primary and secondary outcomes will be.

These choices are then amplified and reflected when evidence from that research is eventually synthesized. In systematic reviews at the other end of the spectrum, the authors have to define their outcomes of interest and accordingly rate the merits of outcomes reported in studies as part of assessing inclusion and exclusion criteria. Preferences concerning suitable outcomes, therefore, have an impact on clinical trial designs and inclusion criteria for systematic reviews.

2.2. Balancing benefits and harms

When investigators synthesize data into systematic reviews and guidelines they attempt to answer the question, ''Does the intervention (as examined across multiple studies) do more good than harm?'' The task of balancing benefits and harms is further complicated by the need to both accurately ascertain the magnitude of effect for a given outcome and estimate the quality of those research results for purposes of weighing. Balancing multiple dimensions of benefit and harm data—relevant outcomes, magnitude of effect, and quality of results—presents a particular challenge in formulating guideline recommendations that often depend on quality evidence synthesis.

Take the example of mammography screening [16]. After lengthy, often politically charged deliberations, the U.S. Preventive Services Task Force concluded that there was an overall benefit to mammography as evidenced by decreased breast-cancer specific mortality despite the harms of false positive test results [17]. On the other hand, a Cochrane review came to the opposite conclusion, citing no decrease in overall mortality after mammography and noted the harms related to increasing surgical interventions [18]. The differences between the U.S. Preventive Services Task Force and the Cochrane review conclusions illustrate differing preferences for certain outcome measures (i.e., VJs) leading to conflicting weighting of benefits and harms of mammography.

2.3. Tolerating uncertainty

The findings of clinical trials as well as systematic reviews always contain some degree of uncertainty due to biasing influences of chance and different types of biases. Accordingly, a third type of VJ that dominates the process of planning a clinical trial also influences the assessment of study quality in systematic reviews—balancing the value of gaining more but less accurate evidence vs. gaining less evidence that is more accurate. VJs are also important in choosing which research designs to take seriously. For example, choosing an experimental versus an observational approach, or choosing important effect modifiers (e.g., ethnicity or social class). Because even the most rigorous study designs cannot eliminate all uncertainty, those evaluating data from individual studies must decide how much uncertainty or potential for bias they are willing to accept when judging the merits of published evidence. These choices are not merely technical in nature but reflect the values of those choosing. Although it is common for investigators to use validated quality assessment instruments, there are no agreed criteria for specifying the weight that should be assigned to limitations in studies' quality once assessed. Unless there are severe concerns, most systematic reviews include all studies in the sensitivity analyses to assess the possible influence of the various quality items [19]. Nevertheless, deciding whether deficits in measured study quality are important enough to exclude the study findings from meta-analysis, for instance, often reflects review authors' tolerance of uncertainty.

Although the often-applied Jadad criteria focus on the randomization process, other review authors prefer criteria that focus on the degree of postrandomization exclusions or baseline imbalances [20]. Recently, for instance, Cochrane reviews of cholinesterase inhibitors in the treatment of patients with Alzheimer's disease concluded an overall benefit for cognitive and global outcome measures after employing the Jadad score for quality appraisal of clinical trials [21–23]. Another systematic review came to a rather pessimistic conclusion after focusing on postrandomization exclusions, baseline imbalances, and a more critical analysis of the reporting of patient randomization [24].

These VJs are similar to the determination of alpha or beta errors in designing studies [10,25]. They can be specified a priori, but the exact levels chosen are determined by social convention or personal preference. We must acknowledge that every choice in this regard requires balancing the uncertainty of being wrong in our inferences about study quality with the probability of missing important signals about true benefits and harms from studies of suboptimal quality. The answer to how much uncertainty in study quality we are willing to accept ought to be dependent on the context (e.g., severity of disease, existence of alternatives) and on the preferences and values of the particular patient population to which the evidence will be applied. These VJs, therefore, have to be informed by knowledge in

statistics and clinical epidemiology as well as by knowledge in medicine and ethics. Because there is no ''one size fits all'' approach for determining how much uncertainty should be tolerated in designing clinical studies or in synthesizing evidence, it becomes important for users of the evidence to be given more information about the investigators tolerance of uncertainty and their rationale for their choices in a given circumstance.

3. Toward greater transparency

In light of the VJs that shape clinical research and evidence synthesis, a major challenge is to provide greater transparency about these VJs. VJs reflect personal or social preferences about which reasonable people can disagree. Because health care decisions are increasingly subject to public scrutiny and collective influences, it is essential that VJs known to influence production, analysis, and synthesis of evidence be identified and disclosed transparently. The principle of transparency should govern the process of evidence production and synthesis because of the public investment in health care and use of evidence to guide decisions in that investment.

Although it might seem impossible to achieve full transparency about all potential VJs in evidence production and synthesis, a degree of greater transparency could be achieved quite easily. For instance, authors of systematic reviews and health technology assessments, the compilers and synthesizers of evidence, should not only state the most important VJs influencing their reviews but also highlight how their VJs differ from the stated or implicit VJs of previously published systematic reviews on the same topic. Disclosing and justifying choices related to the three key VJs outlined here would take modest effort and should improve the social value of evidence produced for health care decision making. For example, technical tools to help decision-makers (including clinicians, policy-makers, researchers, and patients, depending on the context) select among discordant systematic reviews have been published [26]. Explicit knowledge about VJs in the context of discordant systematic reviews might further help to make choices among alternative health care interventions. Over time, disclosing VJs may come to explain some of the otherwise frustrating variation in guidelines, and may empower users of synthesized evidence and guidelines, to ascertain which guidelines make the most sense for their purposes given the stated VJs.

Acknowledgments

This work was completed while Dr. Daniel Strech was a visiting scholar in the Department of Bioethics, National Institutes of Health, MD, USA. We would like to thank Dr. Franklin Miller for his critical review of the manuscript.

References

[1] Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. Ann Intern Med 2001;134:657−62.
[2] Sackett DL, Straus S, Scott Richardson W. Evidence-based medicine. How to practice and teach EBM. London: Churchill Livingstone; 2000.
[3] Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp, et al. Grading quality of evidence and strength of recommendations. BMJ 2004;328(7454):1490.
[4] Guyatt G, Drummond R. Users' guides to the medical literature: essentials of evidence-based clinical practice (Evidence-Based Medicine Working Group & American Medical Association). Chicago, IL: AMA Press; 2002.
[5] Kaptchuk TJ. Effect of interpretive bias on research evidence. BMJ 2003;326(7404):1453−5.
[6] Greenhalgh T, Kostopoulou O, Harries C. Making decisions about benefits and harms of medicines. BMJ 2004;329(7456):47−50.
[7] Feinstein AR. Clinimetrics. New Haven, CT: Yale University Press; 1987.
[8] Lambert MF, Wood J. Incorporating patient preferences into randomized trials. J Clin Epidemiol 2000;53:163−6.
[9] Knottnerus JA, Bouter LM. The ethics of sample size: two-sided testing and one-sided thinking. J Clin Epidemiol 2001;54:109−10.
[10] Upshur RE. The ethics of alpha: reflections on statistics, evidence and values in medicine. Theor Med Bioeth 2001;22(6):565−76.
[11] Molewijk AC, Stiggelbout AM, Otten W, Dupuis HM, Kievit J. Implicit Normativity in evidence-based medicine: a plea for integrated empirical ethics research. Health Care Anal 2003;11:69−92.
[12] Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgments. BMJ 2004;329(7459):224−7.
[13] Tunis S. Reflections On science, judgment, and value in evidence-based decision making: a conversation with David Eddy. Health Affaires 2007;26:w500−15.
[14] Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. The CONSORT Group. Ann Intern Med 2001;134:663−94.
[15] Fletcher RW, Fletcher SW. Clinical epidemiology. The essentials. Baltimore, MD: Lippincott Williams & Wilkins; 2005.
[16] Goodman SN. The mammography dilemma: a crisis for evidence--based medicine? Ann Intern Med 2002;137:363−5.
[17] Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2002;137:347−60.
[18] Gotzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? Lancet 2000;355(9198):129−34.
[19] Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions 4.2.5[updated September 2006], The Cochrane Library, Issue 4. Chichester, UK: John Wiley & Sons Ltd; 2006.
[20] Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996;17:1−12.
[21] Birks J, Grimley Evans J, Iakovidou V, Tsolaki M. Rivastigmine for Alzheimer's disease. Cochrane Database Syst Rev 2000; CD001191.
[22] Birks J, Harvey RJ. Donepezil for dementia due to Alzheimer's disease. Cochrane Database Syst Rev 2006; CD001190.
[23] Olin J, Schneider L. Galantamine for Alzheimer's disease. Cochrane Database Syst Rev 2002; CD001747.
[24] Kaduszkiewicz H, Zimmermann T, Beck-Bornholdt HP, van den Bussche H. Cholinesterase inhibitors for patients with Alzheimer's disease: systematic review of randomised clinical trials. BMJ 2005;331(7512):321−7.
[25] Feinstein AR. *P*-values and confidence intervals: two sides of the same unsatisfactory coin. J Clin Epidemiol 1998;51:355−60.
[26] Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. CMAJ 1997;156(10):1411−6.