The Routledge Companion to Philosophy of Medicine

Miriam Solomon, Jeremy R. Simon, Harold Kincaid

Expert Consensus

Publication details
Miriam Solomon
Published online on: 25 Oct 2016

Accessed on: 08 Aug 2023

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
“Expert consensus” is a shorthand way of referring to the agreement of experts on a matter about which they have expertise. The practice of medicine involves frequent reliance—by clinicians, patients, and others—on expert consensus. Sometimes that consensus is informal (achieved through the usual processes of research and clinical practice) and sometimes formal (developed at a medical consensus conference). An example of an informal consensus is, “The experts (pediatric pulmonologists) agree that early aggressive treatment of cystic fibrosis is best.” An example of a formal consensus is, “The Cystic Fibrosis Foundation recommends that airway clearance therapy be increased as part of the treatment of an acute exacerbation of pulmonary disease.” The goal of this chapter is to investigate the social epistemology of expert consensus in medical contexts. Why do we rely on expert consensus and under what circumstances (if any) is it appropriate to challenge it?

Trust in the knowledge of others is a foundation for all knowledge (Hardwig 1991, Scheman 2001). Our knowledge—whether of auto mechanics, culinary techniques, legal processes, or medical practice—rests on trust in the work of epistemic peers and epistemic experts in various domains. Solipsism (reliance only on one's own mind) is a route to ignorance, especially in scientific and technical fields where individuals are dependent on their education by teachers and mentors, on prior research, and on their collaborators and critics. In medicine specifically, the often-cited reliance “on the literature” is reliance (of researchers and clinicians) on the work of other researchers, and reliability “on medical advice” is reliance (of patients) on the knowledge of medical experts. Expert consensus produces the most trust, since it combines both the epistemic advantages of expertise and the reassurance that several apparently independent individuals have come to the same conclusion. When the experts agree, it is typically thought, claims to knowledge have the most chance of being true, or at least true enough to be useful. Trusting expert consensus seems like a reasonable—although not infallible—strategy.

I find it helpful to think of expert consensus as a modern version of the ancient and medieval “argument from authority.” Instead of reliance on the writing of some individual giant or genius (typically white, male, and deceased), the reliance is on the (hopefully) democratic deliberations of a free (unbiased) research community composed of experts in their fields. For a layperson who knows little about the matter at hand, reliance on expert consensus is an excellent epistemic strategy.

Of course, no epistemic trust, even in the consensus of experts, should be absolute. There was a consensus among astronomers from antiquity until the 16th century that the Earth was the center of the universe, and that turned out to be wrong. There was consensus among surgeons from the late 18th century until at least 1970 that shaving the skin before surgery
prevents infection, and that turned out to be wrong. Experts are not infallible. Nevertheless, the agreement of experts deserves to be taken seriously. It is also the most concise way of appealing to the complex body of evidence and reasoning that experts are assumed to have mastered and based their decisions upon.

Here is a well-known non-medical example of the use of both informal and formal expert consensus in communicating a scientific result. Activist-scholar Naomi Oreskes uses expert consensus to argue for the reality of climate change. She points out that the publications acknowledging climate change vastly outnumber the publications denying it (Oreskes 2004), reflecting an informal consensus. In addition, there has been a formal scientific consensus on the reality of climate change, in the statements of the Intergovernmental Panel on Climate Change (IPCC). These are carefully composed joint statements agreed to by thousands of scientists working on climate change. In medicine, it is similarly common to appeal to expert consensus, either informal (when individual scientists do not attempt to make a joint statement) or formal (as in consensus statements, typically produced at medical consensus conferences).

It is appropriate to interrogate the reliance on expert consensus in some cases by asking questions such as: do any of the experts dissent? If so, what do they say and what do they claim is the evidence for their position(s)? How difficult is it to maintain a dissenting position in this scientific community? Or, to put it another way, what is the strength of peer pressure? Are there assumptions that this scientific community takes for granted that should be challenged? If the consensus is formal, who is included and who is excluded from the process used to make a consensus statement? Leading researchers should be included in some way, but care should be taken to avoid simplistic reliance on markers of prestige such as seniority, institutional affiliation, etc. to identify the positions to be taken seriously. For a sophisticated account of how to ask such questions of an apparent expert consensus, see the work of philosopher of science Helen Longino (1990).

The common medical practice of asking for a second opinion is an example of the reliance on informal expert consensus—albeit a consensus of just two experts. When the second opinion agrees with the first, clinicians and patients feel reassured about the correctness of the recommendation. Sometimes it takes a third or fourth consistent expert opinion—especially when the matter is consequential—to settle any questions. When the expert opinions differ from one another, however, there is much uncertainty about how to proceed. In such cases, decisions are often made based on considerations other than the preference for good health outcomes. For example, they may be made based on cost, and the cheapest intervention selected. In effect, medical experts can lose their voice in practical decision making when they disagree with each other.

Disagreement is not in itself problematic for science—it is a valuable source of constructive criticism and division of epistemic labor (Solomon 2001)—but it is a problem for the role of science in matters of public concern and practical decision making. As cigarette manufacturers and climate change deniers know well, if they can create even the appearance of dissent in science, they can produce public distrust of science (Oreskes & Conway 2010). Similarly, if there is expert disagreement in a medical context, decisions can be hijacked by other interested parties, such as health insurance companies, professional interests, or government policies. Expert consensus is thus important for maintaining the authority of science in the medical context.

The Medical Consensus Conference Movement

While scientific medicine dates back at least to the 19th century, the pace, complexity, and sophistication of medical research increased greatly in the post–World War II era. In the early stages of testing, the effectiveness of most health interventions is uncertain. This is because
the effectiveness tends to be modest and non-uniform: some are helped, whereas others may be not helped or even harmed. The overall effectiveness of health care interventions is reliably discerned through the aggregated results of high-quality clinical trials. By the 1970s, the need arose for authoritative and unbiased evaluations of new medical interventions such as drugs, surgery, and tests. In response, the U.S. National Institutes of Health (NIH) developed an assessment process, based on expert consensus, called the NIH Consensus Development Conference Program. This was the beginning of an international medical consensus conference movement. Begun in 1977, the NIH conferences had a formal structure, taking place over three days, with a written statement presented at a press conference on the third day. Each conference was run like a court of law, with a “judge” (the panel chair), “jury” (10–20 panelists with medical expertise in the area under discussion), and “witnesses” (prominent researchers in the area under discussion). Details of the conduct of the conferences show the effort to give both the appearance and the reality of an objective process, with freedom from bias (whether financial or intellectual), openness to the public for most of the proceedings, and instructions to the scientific “witnesses” to present as they would at a scientific conference rather than as legal advocates. (See Solomon 2015, Chapter 2, for more details about this.)

The expectation was that the medical consensus conference would be a social epistemic process in which experts who might at first disagree with one another come to consensus on the basis of sharing their knowledge of relevant evidence and arguments. It was anticipated that medical consensus conferences would produce authoritative, speedy, and well-publicized statements communicating new and reliable knowledge to clinicians, policy makers, and all interested parties.

Examples of topics of early NIH Consensus Conferences were Breast Cancer Screening (1977), Intraocular Lens Implantation (1979), and Drugs and Insomnia (1983). The immediate reception of the conferences was very positive, both in the U.S. and in other (First World) countries. The NIH model was widely adopted and adapted, so that it became the dominant method of assessment for new medical interventions throughout the 1980s and early 1990s. Examples of U.S. consensus conference programs are run by the Medicare Coverage Advisory Committee and the U.S. Preventative Services Task Force. Examples of non-U.S. programs are run by the Canadian Preventative Services Task Force and the Danish Medical Research Council. A 1990 Institute of Medicine report on the international use of consensus conferences began by stating that “Group judgment methods are perhaps the most widely used means of assessment of medical technologies in many countries” (Baratz, Goodman & Council on Health Care Technology 1990).

There were important differences among the various consensus conference programs. The NIH represented one extreme, focusing on what they called “technical” (meaning “scientific”) rather than “interface” (meaning “broader” issues such as economics, ethics, etc. as well as “scientific”) consensus. The NIH wanted to steer clear of policy issues, whereas other consensus conference programs (especially those in Canada and Europe) eagerly took on the policy questions, even seeing these as their main charge. In Europe, the focus on expertise was modified to allow an emphasis on public participation: the well-known “Danish” consensus conference model, developed at the Danish Board of Technology (now the Danish Board of Technology Foundation) used a panel of selected lay volunteers, along with the guidance of experts. The sense in Europe was that policy deliberations should have—at least in microcosm—the semblance of a public democratic process. In the U.S. and Canada, any public participation was limited to one or two participants on a panel. The European, particularly Scandinavian, method of consensus conferences with public participation coheres with the style of participatory democracy that is politically established in those countries (Horst & Irwin 2010). In North America there has been more deference to expertise.
Although the early reception of medical consensus conferences was largely positive, a minority expressed skepticism about the proposed use of a three-day deliberative process to move from scientific dissent to scientific consensus. For example, Arthur Holleb wrote an editorial deploring the idea of “medicine by committee” and argued that the conclusions reached would be susceptible to bias (1980). Others noted that consensus conferences are not generally used by scientists for their own research purposes. When physicists, or chemists, or biologists disagree about a scientific matter, they do not convene a consensus conference to settle their disagreements; instead, they engage in further research designed to gather new evidence and arguments to support (and/or modify) their positions. This minority expressing skepticism went largely unheard, so great was the early enthusiasm about consensus conferences.

(The case of the Intergovernmental Panel on Climate Change may seem like an exception to the rule that scientists do not settle disagreement among themselves with consensus processes. However, it is not: the IPCC consensus statement is not intended for the scientists, but for the public and governments. It is needed only because of the influence of climate change deniers.)

Over time, experience with medical consensus conferences yielded more nuanced assessments than the early enthusiasm showed. NIH consensus conferences tended to go particularly well when there was already a consensus in the research community, an irony since these conferences were originally designed to develop a consensus that did not yet exist in the research community. When consensus in the research community did not exist, NIH consensus conferences tended to go worse, even on occasion ending with a split panel (Institute of Medicine 1985). Because NIH consensus conferences took more than a year to plan and organizers wanted to wait until there was sufficient evidence to draw plausible scientific conclusions, they usually took place after a consensus was reached in the research community. The NIH argued, plausibly, that they still served a purpose, namely to communicate authoritatively the results of that consensus. (Of course, that purpose might be achieved more easily without such a choreographed public event.)

Medical consensus conferences also had differential impact. In countries with centralized health care provision (Canada and Europe), the consensuses informed health care policy, and thus had greater impact than in the U.S., where health care is more privatized and decentralized. In general, the impact of medical consensus conferences was less than was hoped for. We now know that written dissemination of expert consensus changes clinical practice slowly, if at all. However, in the 1970s and 1980s, this was not appreciated. Instead, efforts went into improving access to the consensus statements (making statements available in pamphlets, in medical journals, and eventually on the internet) rather than into creating more personal networks or incentives for change (strategies that we use now to engineer change in practice).

The Challenge of Evidence-Based Medicine

By 1990, it was generally appreciated that medical consensus conferences can produce an authoritative expert voice, especially when there is already an informal consensus in the research community. Consensus conferences were widely used. The development of evidence-based medicine in the early 1990s, however, destabilized the role of consensus conferences. Evidence-based medicine championed techniques for evaluating complex evidence that were explicit rather than implicit and dependent on expertise in epidemiology and statistics rather than seat-of-the-pants judgment and group process. (See Chapters 18 and 19 on The Randomized Controlled Trial: Internal and External Validity and The Hierarchy of Evidence, Meta-Analysis, and Systematic Review, respectively, for details about evidence-based medicine.)
From the point of view of evidence-based medicine, expert consensus, without an accompanying explicit evaluation of the evidence, is the lowest form of evidence on the evidence hierarchy used in systematic reviews (if it is counted as evidence at all). This is because consensus is susceptible to bias. It can be brought about by non-rational processes such as peer pressure, time pressure, and aggressive leadership. In such cases, active dissent is silenced rather than incorporated into the group deliberative process. Consensus can also be brought about through non-evidential, seemingly rational but actually quite fallible considerations shared by experts, such as pathophysiological reasoning. For example, in the 1990s and early 2000s, there was expert consensus that vertebroplasty was effective for treating pain from osteoporotic spinal fractures. This consensus was reached on the basis of anecdotal reports of pain relief after the surgery, and also on the pathophysiological reasoning that stabilizing the fractures would relieve the pain. In 2009, two randomized controlled trials showed that vertebroplasty is ineffective for treating osteoporotic spinal fractures. Pathophysiological reasoning is quite fallible, and therefore low on the evidence hierarchy. However, it can be persuasive, and is sometimes all that lies behind a consensus of experts.

Thus, it might be expected that with the rise of evidence-based medicine would come the fall of consensus conferences. But that is not exactly what happened, although to be sure the rise of evidence-based medicine forced a reassessment of consensus conferences, a decline in some of the programs, and changes in those programs that continued. Yet expert consensus, and even expert consensus conference programs (sometimes renamed to include the words “evidence-based” along with the term “consensus conference”) have continued to be important. For example, the Medicare Coverage Advisory Committee was renamed the Medicare Evidence Development Coverage Advisory Committee in 2007. The current situation is that expert consensus and medical consensus conferences have a continuing role, integrated with evidence-based medicine, in producing authoritative knowledge. The consensus work typically follows the work of systematic evidence review. There are at least three reasons for this continuation of reliance on expert consensus and consensus conferences.

The first reason is that consensus conferences are often thought of as a check on the prior work of formal evidence review. When well-known experts participate in the consensus conferences, they show their agreement with the evidence synthesis, which may be done by less well-known statisticians. Expert consensus thus adds a layer of assurance about the results of the evidence review. Analyses of the evidence from statisticians and epidemiologists do not stand alone, but are supplemented by a joint statement of credible domain experts in the medical field.

The second reason for the continuing role of medical consensus conferences is that decisions about whether to recommend a medical intervention may rest on policy as well as scientific considerations. Cost considerations, comparisons of harms and benefits, and specific challenges with implementation may need to be considered, often in nationally and culturally specific contexts. Consensus conferences are particularly well suited for considering policy issues in a manner that is perceived as democratic as well as authoritative. The addition of ethics and politics experts to the panel, as well as laypersons (more common in Scandinavian countries), improves the perceived value of the consensus.

Finally, the practice of evidence-based medicine itself rests on a foundation of expert consensus. Evidence-based medicine is really “evidence-hierarchy medicine,” and the exact hierarchy to use is a matter of controversy. Many rating systems are in use. They all rank randomized controlled trials high and anecdotal evidence low, but they disagree on matters such as the status of high-quality observational trials, some regarding them as highest-quality evidence and some not. The GRADE Working Group, established in 2000, is attempting to reach expert consensus on one system of rating the quality and strength of the evidence in clinical trials. It is ironic that expert consensus may supply the foundations for evidence-based medicine.
The Current Role of Consensus and Consensus Conferences in Medicine

The NIH Consensus Development Conference Program, which began the consensus conference movement, was “retired” in 2013 (Solomon 2015). Many of the 1980s and 1990s consensus conference programs, such as the King’s Fund Forum in the UK and the Norwegian Institute for Hospital Research, no longer exist. With the dominance of evidence-based medicine, appeals to “expert consensus” rarely stand alone. But, unlike some other items on the lowest tier of the evidence hierarchy—like anecdotal evidence—expert consensus and even consensus conferences are still very much a part of making knowledge in medicine.

An example of a medical knowledge resource that shows the importance of both evidence and expert consensus is the National Guidelines Clearinghouse. Supported by the U.S. governmental Agency for Health Care Research and Quality, it describes itself as “a public resource for evidence-based clinical practice guidelines.” It has two main requirements for listing clinical guidelines: that they come from a professional group (and in particular, not from individuals) and that they are evidence-based. Professional groups typically devise clinical guidelines in workgroups comprising respected experts who meet face-to-face. The importance of such meetings has not changed, even though the “golden age” of medical consensus conferences is over. Evidence review typically takes place before the consensus meeting, and the results anchor the conference discussions, which focus on clinical practice guidelines.

The evidence review together with an analysis of benefits and harms is, really, all that is needed to evaluate the current state of knowledge. For effective dissemination and uptake of new knowledge by clinicians, the public, and policy makers, the assurance of expert consensus on guidelines based on the evidence review is also needed. The requirement for professional group authorship of the guidelines ensures that experts in the field jointly agree on the guidelines, and thereby increases confidence in them. The guidelines are supported by the reputation of the professional group. Moreover, the participants in the consensus conference begin the process of practice change by taking these results back to their communities, in which they typically play the role of thought leaders. These days, it is understood that publication of guidelines—no matter how trustable the source—is not sufficient for bringing about change in medical practice. Dissemination of recommendations for change is more effective when it is done at the local level, through the active influence of thought leaders and even the creation of local incentives.

Expert Disagreement

So far, I have focused on the rhetorical force created by the dissemination of expert consensus. In some fields—such as cystic fibrosis treatment—there is a high degree of consensus and a correspondingly smooth delivery of effective health care. Atul Gawande (2004) praises the medicine in this field, writing that “cystic fibrosis care works the way we want all of medicine to work.” The flip side is that lack of expert consensus—expert disagreement—leaves a rhetorical vacuum. As I said at the beginning of this chapter, medical experts can lose their voice in practical decision making when they disagree with each other. This creates an opportunity for other interested parties to become more influential, sometimes to the detriment of medical care. The medical profession has resisted this by minimizing their disagreements, as well as by doing “damage control” when disagreement is inescapable.

It is worth looking at an example of such disagreement: the controversy over recommendations for screening mammography, especially for women aged 40–49. I choose this example for two reasons. First, it shows several kinds of effort to minimize dissent. Second, there are important differences between screening mammography and cases when there is consensus on clinical guidelines.
Recommendations for screening mammography have been controversial ever since the technology was developed in the late 1960s. It was certainly expected, in the early days of the technology, that screening mammography would reduce breast cancer deaths, because the prevailing understanding of cancer was that early detection saves lives. The early days of the technology also coincided with the women’s health care movement, and not surprisingly breast cancer screening became both a symbol of women’s self-empowerment and a measure of adequate health care coverage for women’s health. Many professional organizations made early consensus statements about the importance of mammographic screening, yet clinical trials have shown that the benefits of breast cancer screening are at most fairly modest. As the quality of clinical trials has improved, the estimated benefit of breast cancer screening has decreased. Some professional organizations (e.g., the U.S. Preventative Services Task Force and the American College of Physicians) now recommend against routine mammography for women aged 40–49. Others (e.g., the National Cancer Institute and the American College of Radiologists) have stayed with the earlier recommendations of annual mammograms for women in this age group. As this chapter goes to press, the American Cancer Society is in the process of changing its recommendations and taking an intermediate position, making different recommendations for women aged 40–44 and women aged 45–54.

The medical profession has “managed” the dissent over screening mammography (and other controversial interventions) in three different ways—reframing, promoting trust, and producing guideline syntheses. Reframing involves regarding areas of dissent positively, as areas for clinical judgment and patient preferences, rather than as a reason to distrust the experts. Even though the scientific uncertainty is not equal to either patient biological variability or variability in patient preferences, the strategy is a convenient compromise that makes positive use of expert disagreement.

Disagreements about medical efficacy are frequently accompanied by suspicions about costs and health care coverage. It is not unreasonable for someone to suspect that a reduction in the recommended frequency of mammography screening is motivated by a desire to save costs. The medical profession has done much to allay such distrust in the case of screening mammography by insisting that individuals who want more frequent mammographic screening than is recommended should still be covered for this by health care insurance.

Finally, so-called guideline syntheses, produced by The National Guidelines Clearinghouse, are attempts to reconcile different guidelines for the same condition. They are reviews of those of the guidelines that are of high enough quality to be listed (i.e., written in the past five years, evidence-based, and from a professional organization) and a discussion of the areas of agreement and disagreement. In general, such guideline syntheses emphasize the areas of agreement and acknowledge the rationales of the areas of disagreement. Overall, they support medical authority.

Controversy about clinical recommendations is the exception rather than the rule in health care. The case of screening mammography for women aged 40–49 is unlike most areas of clinical medicine, in two ways. First, the evidence is in the “gray area,” with no clear evidence of benefit. For many recommended clinical interventions, the evidence is clearly positive, even if the intervention is less effective than we would like it to be. Second, consensus recommendations for screening mammography were produced before the era of evidence-based medicine. Since the end of the 1990s, it has been general practice for there to be a systematic evidence review before a consensus conference on a recommendation, making it less likely that the consensus conference approach will end up disagreeing with the evidence-based recommendations. This is a structural means of generally avoiding conflict between evidence-based medicine and consensus conference approaches by giving priority to evidence-based medicine.
Conclusions

Expert consensus is an essential part of making medical knowledge. It is vital for establishing the trust needed to disseminate new research findings and clinical recommendations. Medical consensus conferences at first used expert consensus to both evaluate evidence and disseminate conclusions. Evidence-based medicine has taken over much of the former task, but expert consensus is still indispensable for communicating results with authority.

References


Further Reading

