

Cochrane Review as a “Warranting Device” for Reasoning About Health

Sally Jackson¹  · Jodi Schneider² 

Published online: 24 October 2017

© The Author(s) 2017. This article is an open access publication

Abstract Contemporary reasoning about health is infused with the work products of experts, and expert reasoning about health itself is an active site for invention and design. Building on Toulmin’s largely undeveloped ideas on field-dependence, we argue that expert fields can develop new inference rules that, together with the backing they require, become accepted ways of drawing and defending conclusions. The new inference rules themselves function as warrants, and we introduce the term “warranting device” to refer to an assembly of the rule plus whatever material, procedural, and institutional resources are required to assure its dependability. We present a case study on the Cochrane Review, a new method for synthesizing evidence across large numbers of scientific studies. After reviewing the evolution and current structure of the device, we discuss the distinctive kinds of critical questions that may be raised around Cochrane Reviews, both within the expert field and beyond. Although Toulmin’s theory of field-dependence is often criticized for its relativism, we find that, as a matter of practical fact, field-specific warrants do not enjoy immunity from external critique. On the contrary, they can be opened to evaluation and critique from any interested perspective.

Keywords Argumentation · Expert opinion · Health reasoning · Cochrane Review · Toulmin model · Field-dependence · Warranting device

✉ Sally Jackson
sallyj@illinois.edu

Jodi Schneider
jodi@illinois.edu

¹ Department of Communication, University of Illinois at Urbana-Champaign, 702 S. Wright St., Room 3000, Urbana, IL 61801, USA

² School of Information Sciences, University of Illinois at Urbana-Champaign, 501 E. Daniel Street, Champaign, IL 61820, USA

1 Introduction

Argumentation is a natural communication practice that serves, broadly, to test people's beliefs and to reach alignments needed for practical action. This communication practice is not restricted to a static repertoire of moves and message types but constantly changes as people devise new ways to draw conclusions and defend these conclusions to one another. The ubiquitous human practice of seeking advice from experts, for example, has very long historical roots (as pointed out, for example, by Walton 1997, p. xiii), but it is also a basis for decision-making that is in constant flux as experts' own resources for drawing conclusions improve.

This is dramatically displayed by rapid changes occurring in reasoning about health over the past century. During the twentieth century, opinions formed primarily on the basis of doctors' personal experience in clinical practice have lost their former prestige, especially when contradicted by opinions formed through rigorous experimental research and systematic aggregation of experimental evidence. While a non-expert may be unable to judge either the value of clinical experience or the conclusiveness of experimental evidence, anyone can recognize that these are qualitatively different *ways* of reaching conclusions and that a contest between conclusions formed in these two ways will generally favor the latter over the former. Experimental findings can be used to rebut conclusions based only on personal experience much more convincingly than personal experience can be used to rebut conclusions based on experimental evidence. Trust in expert opinion is more or less reasonable depending on how that opinion has been formed.

Our interest in the project reported here is in what can be learned about argumentation generally by looking at the invention and gradual institutionalization of *new* ways of drawing and defending conclusions. The contemporary practice of argumentation is infused with newly invented modes of reasoning that add to, and sometimes transform, humanity's store of resources for reasoning. Increasingly, specialized fields of inquiry undertake intentional design work aimed at improving how they draw and defend conclusions from data (Jackson 2015b). A case study presented below describes the intentional design of the Cochrane Review, one recent invention in reasoning about health. We introduce the notion of a *warranting device* to refer to new modes of reasoning designed, as the Cochrane Review has been designed, to solve the characteristic problems of argumentation within specific fields.

Warranting devices are typically invented to serve the purposes of a specific field. Before formally introducing warranting devices, we briefly review Toulmin's (1950, 1958) work on field-dependent reasoning.

2 Field Dependence in Reasoning and Argument

Field dependence in argumentation theory is the idea that reasoning strategies and standards for evaluation of reasoning might vary from one discourse context to another. Toulmin (1950) introduced this general point in relation to ethical reasoning, observing in the course of his analysis that reasoning and argument serve highly varied purposes in human interaction that affect what kinds of standards define good reasoning in each case. Toulmin believed in field-independent standards as well; field-independent standards that differentiate good reasoning from poor reasoning apply alongside any field-dependent standards, but the latter, he believed, do not travel from field to field. Science, for example, has its own standards for acceptable reasoning and argument, but these do not adequately differentiate between good and poor reasoning about ethics.

In subsequent work, Toulmin (1958) developed the theory of field dependence in a more generalized form, introducing the contentious notion that different fields may employ specialized inference licenses to move from data to claims. The differences between arguments in one field and arguments in another field go beyond differences in the content of premises and conclusions, to include actual forms of reasoning—the warrants that make it possible to point to information as the reason for believing something. Fields, according to Toulmin, can develop forms of reasoning tailored to their subject matter and their goals, and these forms of reasoning might require special standards for evaluation of argumentation within that field that would not make sense elsewhere.

Toulmin’s most widely known contribution to argumentation theory is an abstract (and field-invariant) pattern for “argument layout” that is known as the Toulmin model. The point of the Toulmin model is that an argument is always more than a set of premises and a conclusion—or rather, that “premise,” as understood in logic, actually includes two very different functions, one of which (known as data) is informational and the other of which (known as the warrant) has to do with what links the data to a claim—what allows an arguer to put the data forward as support for a contested claim. Fields may differ in what kinds of claims they make and what kinds of data they adduce. This would be acknowledged by any treatment of logic or reasoning or argumentation, and by itself, it does not entail field-dependent standards of evaluation. A valid logical form like *modus ponens* may be abstracted from arguments of highly disparate content, and this is also true for invalid forms like affirming the consequent or denying the antecedent. Toulmin appeared to have in mind that the inference licenses available to a given field might include not only deeply intuitive forms like *modus ponens*, but also a wide range of other inference licenses, each of which might have some form of special backing that is characteristic of that field. In Toulmin’s own words, “the moment we start asking about the backing which a warrant relies on in each field, great differences begin to appear: the kind of backing we must point to if we are to establish its authority will change greatly as we move from one field of argument to another” (1958, p. 104).

Although the Toulmin model (the D–W–C layout) has become a cornerstone of contemporary argumentation theory, Toulmin’s ideas about field dependence

remain quite controversial within argumentation theory. Toulmin's most explicit work on field-dependent reasoning appears in a text intended for classroom use (Toulmin et al. 1984), and unfortunately, the presentation of key concepts and the examples chosen to teach these concepts are not easy to reconcile with his other works.

One source of dissatisfaction with the notion of field dependence is a lack of clarity in exactly what constitutes a field, both in Toulmin's own writing and in interpretations by others (Zarefsky 1982). Reading Toulmin's various discussions of field dependence, a field might be thought to be a kind of activity (such as litigation) or a domain of knowledge connected with that activity (such as law) or something else entirely. While we cannot resolve the question of what Toulmin himself meant by a field, when we speak of fields, we have in mind something like a professional discipline, with a settled purpose, a body of common knowledge, and a community of practitioners who identify with one another.

Another general point of dissatisfaction, thoroughly addressed by Hitchcock (2003) and by Keith and Beard (2008), is the ambiguity and contradiction in Toulmin's treatment of warrants, problems that have led to pervasive misunderstandings by other scholars. Toulmin described warrants as inference licenses, and Hitchcock (2005, p. 375) clarified that they are not premises (whether stated or unstated), but rules (for which novel justifications may arise) governing the movement from data to conclusions (p. 386). The concept of a warrant has an unavoidable ambiguity arising from the fact that its role in drawing a conclusion is slightly different from its role in explaining how the conclusion was drawn. The term warrant occurs in Toulmin referring both to the way a person draws a conclusion from data and to the way that person describes how the conclusion was drawn. Assuming the truth of what experts say about topics within their expertise is an inference rule; any verbal formulation of how the rule works is a description of the rule, not the rule itself. Toulmin's examples regularly show descriptions of a rule as warrants, and this creates a false impression that all inference rules can be given simple assertive descriptions like "Plants from closely related botanical genera may be expected frequently to contain similar biochemical substances" (Toulmin et al. 1984, p. 336). The inference rules we call warranting devices are better understood as instructions for how to draw an inference—how to work with data of a specific kind to get to a conclusion. This understanding of warrants is closely aligned with Hitchcock's, and our case study below of an invented inference rule will show the particular sense in which a warrant is not a statement that can be made an explicit part of an argument, but truly a *way of inferring* that any statement only describes.

A third objection to Toulmin's ideas is their tendency toward relativism: the idea that field-dependent reasoning can only be evaluated using the standards of the field, which Toulmin might or might not have actually intended. This problem has been much discussed (by Freeman 2005; Bermejo-Lugue 2006, among many others). We will address the bearing of our findings on this issue in the concluding section of this essay.

Perhaps because of the obstacles thrown up by Toulmin's own work, not much actual progress has been made in understanding field-dependent reasoning. For example, a widely accepted representational mechanism known as the Argument

Interchange Format or AIF (Chesñevar et al. 2006) acknowledges the possibility of field dependence by including context in the core model and assuming that context may include domain-specific argumentation rules. But as compared with the domain-independent analysis of schemes, context has yet to be meaningfully elaborated within the AIF. We believe that the missing element in Toulmin’s own understanding of field dependence is an account of reasoning innovations within particular societal contexts—contexts defined jointly by some sort of social purpose, some distinctive subject matter, and some set of mutually-engaged participants. Field dependence, we argue, develops over time through invention of new inference rules, and further, new fields (such as medical science) may form around appropriation of such rules from other fields. We believe that our argument is broadly consistent with Toulmin’s intent, while taking his ideas considerably further than what his own writing on the subject would authorize.

3 Warranting Devices

Expert fields may build up repertoires of reasoning techniques over time, some of which are field-dependent inference rules. They also build up fact-finding strategies and various conventions of form, but these are not our present concern. We will use the term warranting device to describe certain stable inference rules accepted within a given field as dependable methods for drawing and defending new claims within the field’s domain. A warranting device is (1) an inference license (2) invented for a specialized argumentative purpose and (3) backed by institutional, procedural, and material components that provide assurances of the dependability of conclusions generated by the device. Hitchcock (2005) may have had a similar idea in mind when describing “justified warrants” for arguments in specialized fields.

A similar phrase, “argumentative device,” has been used by Mercier and Sperber (2011) and also Mercier (2011) for a different, but relevant, purpose. According to their argumentative theory of reasoning, human reasoning evolved as a device for generating persuasive arguments to be used in attempting to create or maintain agreements among individuals and groups. There is obvious resonance between this view and our view that at some point in human history—not very long ago, in fact—people began deliberately inventing ways to improve on whatever is “hardwired” into the human brain. An evolved inclination to search for winning arguments may well be what drives our inventiveness. Biological evolution and cultural change are difficult to distinguish clearly, but we presume that a human child anywhere on earth is born with the argumentative device as described by Sperber and Mercier; by contrast, the inventions we call warranting devices are culturally transmitted and can change very rapidly.

Warranting devices are designed to meet newly-noticed needs—typically in specialized fields where commonsense reasoning is not enough to answer common classes of questions. When any new warranting device is proposed as a way to draw conclusions from data, other experts may challenge it, pointing out limitations of the device or circumstances in which it will lead to false conclusions (as we will describe later). In the terms of the Toulmin model, critique of the device may lead to

enumeration of possible rebuttals for conclusions warranted by the device. But critique may also lead to adjustments in the device to build in a defeator for a rebuttal. Iterative repair and critique continue, often over long periods of time, until the mode of reasoning itself (not just a particular application of that mode of reasoning) is defeated, abandoned, or stabilized. A stabilized mode of reasoning may come to be taken for granted within a field, rarely questioned and rarely explicitly defended. But before that, any newly proposed mode of reasoning may remain contentious and uncertain for long periods, advocated by some experts and disdained by others. Later in their trajectory, as these strategies gain wide acceptance, they may be taught simply as methods, losing their argumentative provenance entirely. For instance, experimental work in many fields may report use of statistical procedures with no reference of any kind to the body of work that led to the procedures.¹

A warranting device may contain material components that augment human reasoning in various ways and institutional components that underwrite their dependability. For example, a table of random numbers is a material object formerly used to prevent human experimenters from unwittingly biasing their assignments of experimental subjects to treatments; although the tables themselves have been obsoleted by other material objects, random assignment of experimental subjects to treatment conditions has become an institutionalized requirement for making causal claims about human response to treatments. The institutionalization of random assignment of subjects to treatments is too complex to trace here, but it figures, for example, in well-understood requirements for demonstrating the safety and efficacy of any newly developed pharmaceutical. Technical components of a device and justifications for them may be employed in actual arguments long before a stable inference rule emerges as a recognizable mode of reasoning. That is, individual technical components may have a history that predates their absorption into a warranting device.

In certain respects, warranting devices operate like the inference principles of familiar argumentation schemes (e.g., for argument from expert opinion, the assumption that what experts believe may generally be taken as true). Schemes, though, are generally assumed to be domain-independent and stable over long periods of time (Chesñevar et al. 2006, p. 297), while the inventions we call warranting devices are deeply entwined with the state of knowledge in a given domain. For any new warranting device, it will be possible to theorize a set of critical questions analogous to those associated with familiar argumentation schemes; these questions may also go beyond the application of the scheme in a particular case to more abstract issues like what unsuspected limitations the device itself may have. Critical questions needed to evaluate the output of a warranting device need to be discovered separately for each such device, often by seeing how the device fares in actual debate among experts, and then again, in larger contexts (like public debate) where the output of the device may be used as evidence for some further conclusion. Devices may change in response to change in the

¹ Toulmin's description of "warrant-establishing" arguments is relevant, but the inadequacy of his understanding of how the process of establishing a warrant actually works must be discussed elsewhere.

substantive knowledge of the field, as when newly discovered facts about the phenomena being studied expose a previously undetectable way for the device to go wrong. So warranting devices inherit certain features that characterize all human reasoning, but they also introduce new complexities that ordinary schemes do not have.

As we will illustrate shortly, much of the complexity of a warranting device derives from material, procedural, and institutional components that (in Toulmin’s terms) provide its backing. These components are part of the disagreement space around any use of the device, determining the kinds of critical scrutiny an argument generated by the device must withstand. For domains advancing high-stakes claims, like medical research, there are many different motivations for critical scrutiny (scientific commitment to empirical adequacy, pragmatic interest in quality of health care, patient concern for safety, financial interest in health care products, and more), and any of these motivations can lead either to the discovery of new critical questions or to the invention of new strategies for disarming them.

We turn next to more detailed examination of the Cochrane Review, a warranting device of very recent invention. We describe its history in Sect. 4, offer a proposal for how to formally represent the characteristic dependence of any such device on field-specific resources in Sect. 5, and discuss critical questions associated with arguments generated by the device in Sect. 6.

4 Case Study: Cochrane Reviews

We have chosen the Cochrane Review as an initial case for developing the concept of a warranting device. The more general category of systematic reviews has previously been discussed by Hitchcock (2005, p. 386) as a justified warrant for deriving a clinical guideline from a body of research, distinguished from the use of the clinical guideline itself as a justified warrant for treatment recommended to a particular patient. Here we focus on how warrants of this kind come to be justified.

A Cochrane Review is a systematic method for synthesizing a body of medical research for the purpose of informing medical practice. It is a scientific technique, but one that does not involve generating new experimental findings. The aim of a Cochrane Review is to decide what can reasonably be inferred from a body of previously generated findings. Scientific findings come in diverse forms (quantitative/qualitative, experimental/correlational, etc.), but the preferred source of data for a Cochrane Review is a kind of experiment known as a Randomized Controlled Trial (RCT) designed to evaluate quantitatively whether a particular medical treatment is effective or to assess which of several alternative treatments is most effective.

Cochrane Reviews are named for Archie Cochrane, a Scottish doctor and epidemiologist who championed the use of medical experimentation for guidance of clinical practice (Cochrane 1972). Cochrane was not a founder of the Cochrane Collaboration, nor was he an inventor of the device. He died in 1988, about 5 years before the formation of the Cochrane Collaboration that has developed and disseminated the device. Cochrane’s personal contributions to medical reasoning

were of a slightly different kind, relevant to the linking of RCTs to improvements in effectiveness and efficiency of health care.

In the five subsections that follow, we (1) briefly summarize some key technical advances that made the Cochrane Review possible; (2) describe the components of the device in its present state of development; (3) explore the work required to build and maintain the material components of the device; (4) discuss external social pressures on device design; and (5) reflect briefly on the present status of the device in resetting standards for reasoning about health. Case studies of other devices will be needed before the generality of our conclusions can be assessed, but each case study of an important new warranting device has independent value on a par with analysis of an individual argumentation scheme like argument from expert opinion.

4.1 Technical Threads Leading to the Device

Central to the concept of a warranting device is the idea that new inference tools can be invented. Inventions of all kinds typically take advantage of prior work that provides foundational ideas about how some problem might be solved. Medical science has as one of its characteristic inference problems the problem of finding causal relationships between medical interventions and health outcomes. The Cochrane Review is one invention in a series of other inventions that have attempted to solve aspects of this problem. The device combines technical ideas drawn from multiple sources, woven together into a novel way of achieving an objective: the formal aggregation of multiple pieces of scientific evidence into coherent conclusions about causation. Some of these foundational ideas are abstracted from prior methodological inventions within medicine and other scientific fields, while others are inspired by significant technological changes occurring with the rise of computing and information science. Several major technical threads have converged in the design of the Cochrane Review.

Countless large and small inventions over many centuries have contributed to a broadly accepted view of what is required to demonstrate causality in the context of human health and medicine (Bradford Hill 1965). Among these, one of the most important developments is the RCT, adopted within the twentieth century as the preferred form of evidence for claims about the effectiveness of medical interventions, including drugs. The RCT is itself a warranting device, built from a large number of smaller inventions, such as the “control group,” the random assignment of observational units to treatments, double-blind procedures, and others. RCTs appeared in adjacent fields (such as agriculture and psychology) decades before they became common in medical research, but were quickly appropriated into medical science. Boring’s (1954) account of the rise of control groups in biological and psychological research (first appearing in those fields in the late nineteenth and very early twentieth century) explains how the convergence of this innovation with early twentieth century developments in inferential statistics quickly elevated the control group to the status of an evidence standard for all forms of experimental work involving animals (human or otherwise). An important point to notice in the intertwined histories of RCTs and inferential statistics is that innovations in any one field can quickly diffuse into other fields, even if the issues

belonging to the various fields are substantively different. For example, control groups were essential for fields where the experimental treatment involved anything learnable, but once they appeared, they were spontaneously adopted even in fields where experimental subjects could properly act as their own controls. Another familiar example is the spread of “split-plot” designs in agricultural experiments to logically equivalent designs in education and psychology, where what is “split” is something entirely different from a plot of land—such as a class of students—and where an alternative approach might easily have developed based on assignment of many intact classes to each treatment condition.

Completely independently of advances in causal inference, important changes were occurring in the management of print resources: books and journals. Organizing a large library means having some principle for deciding where a given item will be located, so that the item can be found again when wanted. Organizing a literature is a slightly different problem; any given physical collection might include only a portion of the literature, and no one method of physical placement can assure that items sharing an important commonality will be located together. Solutions to this problem began to appear in the nineteenth century, with proposals for classification systems for books as well as proposals for creation of finding aids such as indexes that could allow readers to locate materials through conceptual search rather than through physical browsing of library shelves. By the mid-twentieth century, these finding aids were transitioning from print resources published periodically to electronic resources that were, increasingly, automated. (We discuss one example, an indexing system known originally as MEDLARS, later in this paper.) By the end of the mid-twentieth century, both print and electronic publications were being published with explicit information (keywords and other metadata) included to serve the purpose of indexing.

Both advances in causal inference and advances in the management of literature are needed to account for the emergence of a new scientific practice known as meta-analysis. With appropriately conducted experiments accumulating rapidly on many specific research topics, it became obvious that drawing conclusions about these topics meant looking not at individual research results but at bodies of work (at least partially) identifiable from indexes. In fits and starts, scholars in varied fields tried various strategies for research synthesis, including just tallying up the number of experiments supporting or failing to support a given hypothesis (later pejoratively described as the vote-counting method). But no later than mid-century, dissatisfaction with these methods prompted serious theoretical work on combining statistical information. (See, e.g., the informal histories offered by Glass 1976; Rosenthal 1984).

By the 1970s, the new methods proposed for statistical aggregation had become known collectively as meta-analysis. These methods were energetically advocated by a small number of behavioral scientists—and greeted with great suspicion by a much larger number of their colleagues. Motivated by the skepticism with which these new methods were regarded, Cooper and Rosenthal (1980) pressed the case for meta-analysis by conducting an experiment in which qualified reviewers were given a stack of studies and either instructed to review them using customary narrative procedures or to review them using supplied meta-analysis procedures.

Those using meta-analysis procedures were, according to Cooper and Rosenthal's interpretation, better able to judge the strength of evidence contained in the set of studies (less likely to see the studies as inconclusive). Although the Cooper and Rosenthal study does not provide particularly strong evidence for the validity of the meta-analysis procedures used at the time, it served the important rhetorical function of exposing unmistakable weaknesses in the narrative and interpretive methods that were, before meta-analysis, the state of the art for aggregation of scientific findings.

Meta-analysis had such pronounced argumentative advantages over narrative reviews that attention quickly shifted away from critiquing the core ideas of meta-analysis, and toward active effort to improve the practice of meta-analysis by building a body of technique and assembling associated resources. An important additional detail is that the rise of meta-analysis as a tool for synthesis fed back into practices of primary researchers. Since the value of meta-analysis is greatly amplified when primary research is conducted with meta-analysis in mind, editorial policies began shifting toward requiring the reporting of statistical information needed for later cross-study comparison. Within a surprisingly short time, meta-analysis became the preferred method for reviewing empirical research for a number of fields, including education, psychology, communication, and other social sciences, rapidly improving and stabilizing its procedures through pre- and post-publication peer review. (For a sense of the discourse surrounding the development and justification of these procedures, see Zwahlen et al. (2008) and Hedges (1986); there are many other such articles in other fields where meta-analysis has been appropriated.) The rise of meta-analysis did not just alter the way research synthesis is conducted, but also exposed facts about variability affecting the interpretation of individual studies (O'Keefe 1999).

Relatively late in this game, in 1989, a major 2-volume synthesis of research on pregnancy and childbirth appeared (Chalmers et al. 1989), with a foreword written by Cochrane praising the work as "a real milestone in the history of randomised trials and in the evaluation of care." This was the first major systematic review in health science, a massive undertaking involving 10 years of effort to review over 3000 controlled trials published since 1950 (Young 1990). Very soon thereafter, in 1993, the Cochrane Collaboration (now known simply as Cochrane) was formed to support the production of similar reviews across a wide range of medical topics (see Bero and Rennie 1995 for a contemporaneous account), integrating the quantitative methods of meta-analysis wherever possible with a body of technique for locating all relevant evidence within a large and diffuse literature.

Considering Cochrane Review as an invention—that is, as a technical achievement—we can trace a large number of prior achievements that made this invention feasible. These include closely related advances in causal inference and proposed improvements in evidence synthesis, but also completely unrelated advances in humanity's ability to manage an ever more massive legacy of prior writing.

At present, the work of Cochrane includes not just the production of reviews, but also the development of standards for proper conduct of the reviewing work, coordination of information resources, methodological innovation, and more. Although not all Cochrane Reviews employ meta-analysis, the device itself is

designed to avoid the problems of traditional reviews that were so clearly exposed by meta-analysis, as we explain next.

4.2 Components of the Device

A Cochrane Review is a synthesis of evidence conducted using very well-defined procedures outlined in an official handbook (Higgins and Green 2011). These reviews assemble evidence that already exists in a clinically-relevant scientific literature, typically from RCTs of health interventions. The input to the review consists of evidence that nonspecialists (including journalists) would very likely consider to be inconclusive or even inconsistent—typically, a large number of individual studies whose separate conclusions about the effect of an intervention vary in size and even in direction of the effect. For an expert, the evidence, while variable, does not appear inconsistent. A Cochrane Review treats study-to-study variation in findings from multiple RCTs as normal and unremarkable, and reviewers draw inferences from this evidence in a highly disciplined way.

The Cochrane Review has already achieved the status of a trusted warranting device, largely because its procedures are so explicitly linked to critical questions on which earlier styles of research synthesis regularly failed. These procedures include exhaustive search for relevant studies; use of scoring rubrics for evaluation of the relevance and strength of evidence in each study; prescribed methods for combining information quantitatively; preferred methods for presentation of findings; and more. Each of these procedures addresses possible vulnerabilities in any review’s synthesis of evidence.

For instance, the exhaustive search and the requirement to include all discoverable relevant evidence are defense against any charge of cherry picking, even though it is understood that no method will guarantee capture of all potentially relevant references (Aagaard et al. 2016). In combination with material resources to be described in Sect. 4.3, the methodical search procedures required for a Cochrane Review make it hard for a critic to object that evidence was assembled to fit the reviewer’s own hypothesis. Reviewer bias is further minimized through highly structured procedures, defined in the Cochrane official handbook (Higgins and Green 2011). Before conducting a review, the reviewing protocol is first defined, based on standard methods. Reviews are required to have standard contents in pre-specified categories, and they must follow the handbook’s guidelines for the data and analyses. Reporting is further standardized by the use of a suite of software tools that must be used in authoring Cochrane Reviews, including templates for report generation.² Cochrane reviews cannot be sponsored by commercial sources that have an interest in the outcomes of a review, and authors’ conflicts of interest, including work on studies that are synthesized, must be declared (Higgins and Green 2011, Sect. 2.6).

² Review Manager (RevMan) is a software tool distributed by Cochrane as research infrastructure. The current version, RevMan 5.3, is available for downloading from <https://web.archive.org/web/20170713040613/http://community.cochrane.org/tools/review-production-tools/revman-5>.

Counter-arguing individual studies (a once-common practice in narrative reviews of literature) is replaced with careful and explicit coding decisions applied impartially to the entire corpus of potentially relevant studies. The use of scoring rubrics for evaluation of the relevance and strength of evidence in each study ensures that researchers apply the same judgmental criteria to each study, rather than scrutinizing some results very critically while accepting others without scrutiny.

Against a charge that a synthesis of research is only as good as the body of primary research available for aggregation, the Cochrane community (said at present to include more than 37,000 contributors from over 130 countries) has adopted a formal practice of “grading” the strength of the evidence base itself (Guyatt et al. 2011; Balshem et al. 2011), to reduce the risk of implying that the conclusion best supported by a current body of evidence is also, on its own merits, a strong and dependable conclusion.

Prescribed methods for combining information quantitatively, when appropriate, ensure that all evidence is taken into account in a consistent manner. Meta-analysis can be used if there are a sufficient number of studies estimating the same effect—using designs similar enough to allow for consistent measurement of effect size. The standard way to describe these study results is a “forest plot” that allows readers to inspect results on a study-by-study basis (Higgins and Green 2011, Sect. 11.1). Figure 1 is an example of a forest plot, illustrating that although the information contained is quite technical, the visual display makes the results intelligible even to motivated non-experts. Specifically, the plot shows at a glance that studies in this review are not completely uniform in their results, and at the same time shows that the treatment (a form of sex education aimed only at promoting sexual abstinence) produces worse outcomes than control conditions in most of the experimental comparisons. Forest plots make it easy to spot outliers in a set of experiments, and they make it hard for anyone to push strong claims based on a single experiment that sits at either extreme of the distribution of results.

Systematic review methods are becoming trusted inferential tools, but they are still in a period of rapid methodological innovation, and this is likely to continue for some time. As these methods gain credibility among experts, additional changes may occur in the practice of primary research as researchers attempt to anticipate the use of their primary reports in various forms of aggregation. Other related changes may occur in the standards editors and article referees apply during prepublication peer review.

4.3 Construction of Material Components for this Device

The ability of a warranting device to function as a dependable inference rule may rest on material components that have to be assembled on purpose to support the rule. This is certainly true of the Cochrane Review. The most important material components of the Cochrane Review are large curated collections of prior work available to reviewers. Two databases—MEDLINE and CENTRAL—merit further examination as technological innovations that have themselves been created through extensive efforts to curate literature.

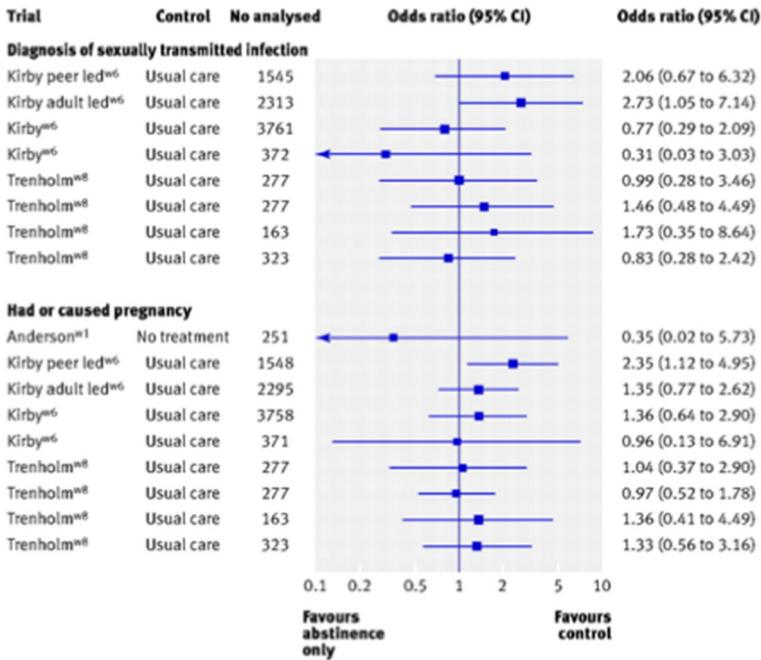


Fig. 1 Reproduced from BMJ, Underhill, Montgomery, and Operario, 335 (7613):248–259, (2007) with permission from BMJ Publishing Group Ltd

Procedures for managing and documenting the retrieval of prior work have become increasingly detailed and documented over time (Lefebvre et al. 2013), and have developed into a chapter of the Cochrane Handbook (Lefebvre et al. 2011) called “Searching for Studies” that is under the stewardship of the Cochrane Information Retrieval Methods Group. The chapter provides basic information about what to search for, the importance of searching in multiple sources, and search strategies and filters³ appropriate for the most common databases. Above all, authors are advised to consult a “Trials Search Co-ordinator” (the information specialist associated with their Cochrane Review Group) and/or a local health librarian early in the process. Computer-based search has greatly facilitated Cochrane Reviews, but even so, review authors are admonished to search in multiple sources, because no retrospectively constructed database can guarantee comprehensiveness. Therefore, reviewers are also expected to search a variety of sources, including MEDLINE, EMBASE, CENTRAL and the review group’s Specialised Register, to identify every possible relevant item, and to examine each item for whether it meets inclusion criteria. A typical Cochrane Review will identify thousands of potentially relevant items and winnow these to a few dozen studies that actually provide relevant data on the question the review is designed to answer.

These procedures assume reliance on material resources, some created by Cochrane, and some created and maintained by other trusted parties. MEDLINE, for

³ Many developed and reviewed by the UK InterTASC Information Specialists’ Sub-Group.

instance, is a selective index to the medical literature that was developed as a byproduct of the MEDLARS project, to more efficiently produce the Index Medicus (a printed periodical started in 1879). Computer typesetting of this monthly printed guide at the US National Library of Medicine gradually changed the way the literature could be searched. Starting in 1964, searchers could request information by telephone, mail, or in-person visit; “trained search analysts would access the system for the designated information,” and the requestor could expect a bibliography in 3–6 weeks (Office of Technology Assessment 1982, p. 19). Access methods have varied as information technology evolved,⁴ and eventually end-users were able to conduct searches without the help of intermediaries. Today any Web user can search MEDLINE online, or download selected search results or even the entire contents of the database. A selective database of high-quality resources, MEDLINE’s contents have changed over time: A committee determines which journals to index, and journals can be removed as well as added. Retrospective data loads and digitization have added some records even from before 1964, and backfiles are no longer searched separately. The rate of change has also varied: Starting in June 2014, new citations could be added to MEDLINE 7 days a week (U.S. National Library of Medicine MEDLINE FactSheet).

CENTRAL, the Cochrane Central Registry of Controlled Trials, was created in 1993 because of a key problem with MEDLINE: Reports of RCTs could not be systematically identified by searching the database. In fact, one study found that about half of the available trials would be missed if MEDLINE were the only source searched, even if they were contained within the MEDLINE collection, because the indexing did not include any code to distinguish trials from other kinds of studies (Dickersin et al. 1994). Initially, a collaboration between Cochrane and the National Institutes of Health was launched to improve MEDLINE indexing, by tagging two Publication Types: RCTs and also Controlled Clinical Trials—trials that may have been randomized but were not explicitly described as such (Dickersin et al. 2002; Harlan 1993; Lefebvre et al. 2013). Cochrane’s carefully constructed search filters helped winnow likely RCTs from electronic searching (for an example, see the Appendix of Dickersin et al. 2002). In addition to electronic searching, individual Cochrane members used handsearching (page by page manual examination of journals and conference abstracts) to identify mentions of RCTs even in items not then typically indexed in MEDLINE.⁵ Cochrane’s wide geographic range facilitated extensive checking of non-English language sources. All of these materials were later used in CENTRAL, along with records culled from Elsevier’s EMBASE database (Dickersin et al. 2002).

Maintenance of CENTRAL is ongoing. Each month, new records are added, drawing on systematic searches of MEDLINE and EMBASE, handsearches of approximately 2400 journals, as well as materials added to the Specialised Registers maintained by over 50 Cochrane Review Groups (Cochrane Library, CENTRAL

⁴ Including paid “on-line” telecommunications (Office of Technology Assessment 1982, p. 22), CD-ROM systems, and “Grateful Med”, a PC-based front-end retired in 1991 (Contaxis 2016).

⁵ Handsearching started from 1948, the publication date for a landmark RCT (Medical Research Council 1948).

creation details). To aid in the time-consuming task of screening database records, in 2014 Cochrane introduced a citizen science project called Cochrane Crowd (Cochrane Crowd). Anyone can sign up for the RCT classification task. After completing a 20-item training set, volunteers are presented with titles and abstracts to classify as ‘RCT/CCT’, ‘Reject’, or ‘Unsure,’ and responses across volunteers are aggregated (Cochrane Crowd; Noel-Storr et al. 2015). Disagreement (which occurs for only 6% of titles presented to volunteers) escalates the case to a more experienced ‘resolver’; otherwise materials are directly added to CENTRAL (or discarded) once three volunteers agree. Validation studies (Noel-Storr et al. 2015) have found that the crowdsourcing procedures, including the escalation procedures for cases with disagreements, result in over 99% accuracy, as compared with the normal procedure previously followed (which used the reconciled judgments of pairs of Cochrane experts).

The point of all this effort is to provide *in advance* the strongest possible assurance, for any individual review, that nothing has been overlooked due to carelessness or personal bias. Instead of leaving the thoroughness of a search to the ingenuity and perseverance of individual searchers, the expert community as a whole invests in creating a repeatable and accountable method that can be presumed to result in as complete a collection of evidence as possible. Of course it is still possible for a search to be incomplete, but the fact that reviewers report exact details of search procedures (including the exact query strings used) means that any objection to the completeness of a search would also need to specify what more could have been done (for example, by showing that additional query strings returned relevant items that the original strings did not, or by showing that the database itself systematically excluded relevant items). Of special interest here for understanding warranting devices is the mobilization of a field’s effort around material requirements for the production of strong arguments. Because these material resources exist apart from any one context of use, they are far less subject to challenge on grounds of reviewer bias in the search for relevant evidence.

Although the Cochrane Review is relatively stabilized in the sense that it has become a trusted way of arriving at conclusions, its form is by no means static. On the contrary, Cochrane has 17 methods groups, charged with addressing various ways of strengthening the device. These groups are organized around different types of topics, including disciplines (information retrieval, statistics), sources of evidence (qualitative evidence, non-randomized studies, individual participant data, prognosis studies, diagnostic tests, patient reported outcomes), evidence assessments (grading evidence, risk of bias), and policy applications (priority setting, economics, equity). The remaining groups are formed around kinds of evidence synthesis: three concern types of meta-analysis (prospective meta-analysis, individual participant data meta-analysis, and network meta-analysis).

In evaluating the credibility of a warranting device like the Cochrane Review, the due diligence exercised by expert practitioners is of central and fundamental importance. This is especially so when non-experts must decide whether or not to trust the conclusions of experts (Jackson 2015a). We return to this point in Sects. 6 and 7.

4.4 Social Factors in Device Development

Warranting devices change mainly to overcome discovered weaknesses in the conclusions they support, but they may also change for other reasons, such as making an enterprise more efficient overall. This echoes a familiar finding in science and technology studies (following Pinch and Bijker 1984) that technologies do not always develop in such a way as to prefer the superior technical option, but often choose options that balance technical superiority against other values. Cochrane Reviews are but one style of research synthesis, and they compete with other technological concepts (including meta-analysis and narrative reviews). Reviews have been considered an ever evolving ‘family’ (Moher et al. 2015) comprising numerous categories (Grant and Booth 2009). Despite the unquestioned rigor of the Cochrane methods, a Cochrane Review must compete (for expert adherents and for policy consumers) against other types of evidence synthesis, including other types of review.

One important challenge is how to do more in less time. Conducting a Cochrane Review is a labor-intensive process, typically taking a team of reviewers 1 to 2 years or more. Cochrane has formed a working group (the Cochrane Rapid Reviews Methods Group, formally established in October 2015; see Garritty et al. 2016) to develop methods for answering questions more quickly and better meeting policymakers’ needs, while maintaining Cochrane’s rigorous standards. Compared to systematic reviews, rapid reviews (RR) are faster to conduct (under 6 months and perhaps just weeks rather than years; see Khangura et al. 2012). At least 29 international organizations conduct rapid reviews, but there is no standard approach, and there is limited agreement as to which standardized methods should apply to rapid reviews (Polisena et al. 2015). As the Cochrane Rapid Reviews Methods Group has noted, “While RR producers must answer the time-sensitive needs of the healthcare decision-makers they serve, they must simultaneously ensure that the scientific imperative of methodological rigor is satisfied” (Garritty et al. 2016).

The importance of noting these non-logical and non-epistemic factors in device development is to acknowledge that any warranting device, being a human invention, may find itself in competition with other proposed warranting devices. Often, warranting devices take shape around multiple competing goals, sometimes involving compromise, for example trading off timeliness against tightness of argument. Both the strategies used to build confidence in a device and those used to evaluate acceptable tradeoffs between rigor and efficiency may provide insight into “warrant-establishing arguments” (anticipated but not adequately theorized by Toulmin).

4.5 Current Status of the Device

The work invested in making Cochrane Reviews more credible has been immense. It has involved not only accumulation of vast collections of scientific reports, but also production of metadata, development of new annotation systems, invention of search tools and strategies, and much more. This device replaces (and obsolesces) styles of literature review that were common until just a few decades ago—one-off

arguments about a body of literature whose credibility was nearly always tied to the personal credibility of the individual reviewer. Perhaps most intriguing is how, in changing the way a community reasons with evidence, the device also shapes how new experimental evidence itself gets produced, presented, and assessed.

5 Representing Arguments Warranted by Devices

In a very preliminary way, we want to consider the challenges of including warranting devices like these in explicit models of argument. One initially plausible way to think about warranting devices is as extensions of the set of schemes available to reasoners: Some warranting devices improve on familiar reasoning patterns, while others allow for making sense of data that previously would have been regarded as uninterpretable, inconclusive, or inconsistent. Using the AIF ontology (Chesñevar et al. 2006), warranting devices would be better represented within argument networks as scheme nodes than as information nodes. Like any scheme, a generally-accepted warranting device can be deployed in deriving a claim from a set of data, and it can also be invoked in justifying the claim or explaining its derivation.

Despite this similarity in function between devices and schemes, it does not seem promising to try to theorize warranting devices as subclasses of familiar schemes. For example, it is tempting to try to account for a device like the Cochrane Review as a variation on argument from expert opinion. Wagemans (2014, p. 52) divides argumentation from expert opinion into “argumentation from professional expert opinion” and “argumentation from experiential expert opinion,” and arguments tied to particular devices would (presumably) be variations within the “professional” branch. What makes this theoretical move attractive is that it would allow different kinds of expert reasoning to inherit critical questions from parent categories while adding new critical questions tailored to each newly identified device. But there are obvious disadvantages to this approach as well. The most serious of these is that it ignores how the experts themselves reason, and focuses only on what happens in the “second-order predications” (Wagemans 2016b) that are the defining feature of arguments from expert opinion. But how the experts themselves reason also needs theorizing; before anything can figure in an argument from expert opinion, at least one expert must have done some reasoning, and there should be some way to connect questioning of the second-order predications with questions that could be raised directly about the first-order predications. Questions that experts raise in arguments with other experts are not the same as questions non-experts raise about appeal to expertise, but neither are they entirely unrelated. For example, the harder it becomes for a non-expert to evaluate the details of the experts’ reasoning, the more important it becomes to evaluate evidence of due diligence by the experts (to know what critical questions the experts themselves have considered). At a minimum, the built-out nature of a device (for example, its reliance on material components such as those underwriting Cochrane Reviews) needs some presence in both sets of critical questions.

A more promising path is to focus on arguments that *use* the warranting devices and try to understand these arguments before thinking about how their conclusions might function as grounds for further argument. Here we begin with Toulmin's insight that the backing for a warrant is what most clearly exhibits field dependence. Warranting devices include various kinds of assurances offered by the field. The novel modes of reasoning that we call warranting devices are best understood as warrants (inference rules), *together* with their backing. Building on Toulmin's hints, we believe that the substantive elements of backing for a warrant may be intellectual assets under the stewardship of a particular field, and this is part of the meaning of field dependence.

But the application of Toulmin's familiar D–W–C layout is far from straightforward. A Cochrane Review (meaning the published report, not the reviewing work on which it is based) is a very complicated text, prepared using a standardized reporting template, and with conclusions of several different kinds. A typical Cochrane Review will advance one or more conclusions about effects of a medical intervention, a further conclusion about what medical guideline is most consistent with the intervention's effects, and possibly a conclusion about the overall condition of the research area within which the review is conducted. Each of these classes of conclusions is developed using Cochrane guidelines and acquires credibility not only from what the reviewers themselves contribute but also from the institutional, material, and procedural assurances provided by the Cochrane organization. Other kinds of reviews, however carefully performed, do not automatically benefit from these assurances.

A practical example demonstrates the complexities of representing the kinds of arguments that appear in Cochrane Reviews. Demicheli et al. (2012) published a Cochrane Review on the efficacy and “adverse effects” of the combined measles-mumps-rubella vaccine (MMR) that is administered to young children around the world. This review is particularly worth study because of its relevance to an ongoing public health controversy, over whether parents should agree to vaccination of their children and (in the US) over whether they should even have the choice of refusing vaccination. Scientifically unsubstantiated belief that MMR may cause autism has been circulating for decades, and consequently, this review has been of interest not only to the expert community but also to the public. The review has received significant attention not only in news reporting but also in social media; its Altmetrics “attention score” identifies it as among the most cited scientific resources of its age and type.⁶

The review presents conclusions in three formats: (1) in a detailed report that resembles any other scientific research report, with full explanations of methods and findings; (2) in an abstract that summarizes significant content of the full report; and (3) in a “plain language summary” that attempts to provide a brief statement of the study's conclusions that is both accurate and intelligible to a general non-scientific audience. Each of these presentations contains threads having to do with the efficacy of the vaccine (with how effective it is in preventing three diseases) and threads having to do with the safety of the vaccine (or rather, with the adverse effects that

⁶ <https://wiley.altmetric.com/details/610980>.

might be associated with administration of the vaccine). This review contains a large number of empirical claims about many different adverse effects, each supported by a subset of all the data available to the reviewers. It also contains advice as a separate claim (from the Author’s Conclusions section): “Existing evidence on the safety and effectiveness of MMR vaccine supports current policies of mass immunisation aimed at global measles eradication and in order to reduce morbidity and mortality associated with mumps and rubella.”

We select for detailed discussion one empirical conclusion because of its many mentions in news and social media: the claim that “no significant association could be assessed between MMR immunisation and ... autism ...” (as formulated in the Discussion section) or that “Exposure to the MMR vaccine was unlikely to be associated with autism ...” as formulated in the Abstract). In both passages autism appears in a list of other conditions evaluated as possible adverse effects of vaccination. But the clear intent of both formulations, especially given other passages in the detailed discussion of autism as an outcome, is to say that *there is no evidence linking MMR to autism*.

Applying the Toulmin model to expose the reasoning behind this conclusion requires identification of the data put forward in support of the conclusion as well as a warrant to connect the data and conclusion. Several candidates for “data” are available: the entire body of primary research that was available to the reviewers, the narrow collection of primary research meeting criteria for inclusion in the review (a total of 54 studies), or the still narrower set of studies from this collection that actually contain evidence related to vaccination and autism (a total of 10 studies that included autism as a variable of interest). Although a case could be made for each of these choices, the most natural and obvious choice is the third: the set of studies that actually report findings on vaccination and autism. These 10 studies are listed, with classification data of various kinds, in Table 9 of the review, headed “MMR and Autism.” Besides information on methods used and other study characteristics, the table also provides reviewers’ judgments of study quality (as “Risk of Bias”) and a summary statement of the study’s results. For example, the results for one study are given as: “No temporal association between onset of autism within 12 months (RI 0.94; 95% CI from 0.60 to 1.47) or 24 months from MMR vaccination (RI 1.09; 95% CI from 0.79 to 1.52).” The most direct basis from which the conclusion is drawn is the list of results as given in the table. Each result, being the reviewers’ assertion about what a corresponding primary study shows, could be questioned critically and unfolded into a defense of the reviewers’ interpretations, so the data are backed by other information that could, in principle, require examination. In addition to these ten result statements, one additional statement is needed because of the particular nature of the claim: that there are no other discoverable studies that report relevant evidence on the relationship between MMR and autism.

Our analytic focus is the warrant that connects the data (ten results and an assumption that no others exist) and the conclusion (that no evidence supports a link between MMR and autism). Presented with a pile of the ten studies, with or without assurance that these exhaust the relevant evidence, a person can draw a conclusion in many different ways. For example, one might draw one’s conclusion from such a

pile of studies by choosing the largest, newest, or most rigorous of the available studies and formulating a conclusion consistent with its results. Or if the individual studies support contradictory conclusions, one might count the number leaning one way and the number leaning the other way, drawing one's own conclusion to conform with the majority. These obviously inferior possibilities are meant only to show that the warrant needed is some kind of *conclusion-drawing rule* (to introduce a useful paraphrase of inference rule). An important point to notice is that different conclusion-drawing rules can, in principle, lead to different substantive conclusions. This is what Cooper and Rosenthal (1980) tried to show in the study discussed earlier by having their experimental subjects draw conclusions from studies using either commonsense reasoning or formal meta-analytic methods. The conclusion-drawing rule that Demicheli et al. followed is one of several options laid out in the procedures required for a Cochrane Review. The rule, together with its various forms of backing, is what we call the warranting device.

Figure 2 presents a proposed analysis of the argument for this particular claim, that “There is no evidence linking MMR to autism.” The point of a Toulmin diagram is to expose the strengths and weaknesses of the reasoning behind a conclusion, or to put it in slightly different terms, to expose the disagreement space around the conclusion. Disagreeing with any element of the data is possible in principle, and leads to discussion of what the primary study did or did not show. Rejecting the warrant (for example, by putting forward an opposite conclusion based on a different conclusion-drawing rule) leads discussion into the backing that justifies the choice of conclusion-drawing rule. The backing for the Cochrane rule contains assurances (provided by the Cochrane community rather than by the individual reviewer) that the conclusion-drawing rule employed by the reviewer is dependable. A thought experiment can clarify the role of backing: If two different possible warrants (two different inference rules) were to produce different conclusions from the same data, the reasons for trusting one rather than the other would need exploration, and the content of this exploration is the backing for each of the alternatives' warrants. As with any other inference rule, an invented warrant may also have its record of success and failure as part of its backing.

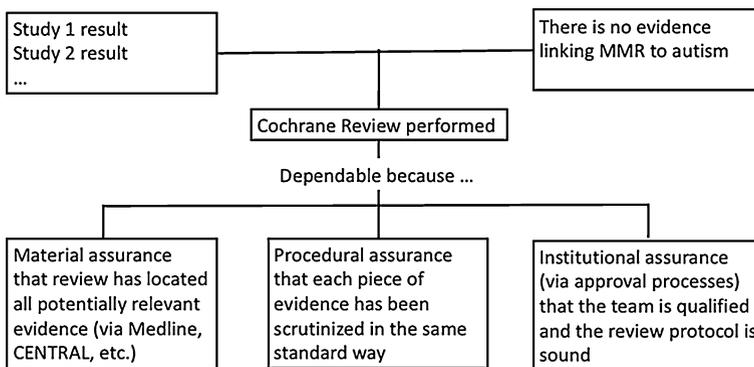


Fig. 2 A Toulmin diagram of one claim from Demicheli et al. (2012)

Note that none of the backing elements, nor the inference rule itself, are explicitly mentioned in any of the presentations of the argument that “There is no evidence linking MMR to autism.” All of this is communicated by identifying the work as a Cochrane Review. In general, a warranting device will allow an argument to be presented without explicit reference to all of the elements of the backing, even though each element of backing is in some sense part of the disagreement space around any individual claim generated by the device and in some sense implicit in the application of the inference rule. Recovering the inference rule and the elements of backing normally requires going “outside the text” to find the rationale for the device and the details of its design. For Cochrane Reviews, these are extensively documented in the Handbook.

An interesting complication in modeling field-dependent reasoning is that, if it is truly field-dependent, the reasoning will be less transparent to those outside the field than to those inside. Once a device has become stabilized within a field, experts accept conclusions generated by the device unless there is some specific reason to object to the particular use of the device (for example, an objection based on incompleteness due to poorly chosen query strings), simply taking for granted the dependability of the device itself. But from outside the field, the device (including the assurances that back the inference rule) may still need defense. A non-expert may question why experts come to the conclusion they do or may question whether experts’ acceptance of conclusions generated by the device are also worthy of acceptance by others who do not share the experts’ common interest and background assumptions. Both kinds of questioning can lead down to examination of the device itself, first to the inference rule and then still deeper to its assurances. A distinct advantage of the Toulmin model is its consistency with the fact that anyone (expert or non-expert) who questions the validity of a conclusion warranted by a device can, in principle, explore any element of the argument for the conclusion, drilling down as far as necessary to satisfy one’s own doubts.

In our proposed modeling of warranting devices, we take the primary purpose to be exposing the characteristic disagreement space around a conclusion generated by the device—the avenues along which criticism of the conclusion might travel, some of which are navigable only from inside the field and some of which are navigable only from outside the field. The analytic decision to include a particular element in the backing is based on whether a critical question formulated about that element would be relevant to the credibility of the conclusion.

A warranting device gains its status over time through incorporation of various assurances of its own ability to deliver reliable conclusions, potentially including new field-specific resources that underwrite the device as a whole. “Assurance” should not be taken as any absolute guarantee of the correctness of conclusions drawn using a device. An assurance is a kind of responsibility, often assumed by the field as a whole rather than by the individual reasoner, for the overall performance of the device (relative to what else might be done to underwrite claims of the type for which they are designed). Assurances can take many forms, and we are not prepared as yet to try to provide any sort of classification, except for the current case study.

For the Cochrane device, a complex system of assurances has already emerged, including institutional, procedural and material resources that are managed by the field as a whole so as to underwrite the dependability of the device. For example, Cochrane Reviews require exhaustive search for relevant evidence, and a reasonable assurance of an exhaustive search is a repeatable method for finding every potentially relevant bit of existing medical research on a given topic. This depends (in practical terms) on gathering and cataloguing all medical research, without regard to topic, and then locating those few dozen studies (among hundreds of thousands) that might address the question of interest. Leaving this responsibility to the individual reviewer means leaving the conclusion open to charges of selectivity, bias, and even sloth; a Cochrane Review has a certain protection from any such charge, because the reviewer delegates responsibility for the adequacy of the search to a resource that is as complete as the field as a whole has been able to make it. The exhaustiveness of a literature search is assured (in the special sense above) by the existence of databases that can be defended as having left out nothing relevant and by full disclosure of query methods. It is still possible to raise questions about whether a given reviewer has conducted a diligent enough search for all relevant evidence, but now that extensive investment has been made in creating searchable databases, the burden of proof has shifted to anyone who wants to say that the search was not thorough enough. So one assurance, for the Cochrane device, is the existence of a collection of research reports that have been curated on behalf of the field as a whole. Besides exhaustive search, a review should include evidence from every relevant report (that is, should not simply discard inconvenient evidence). Explaining every exclusion from analysis is a procedural assurance, and this too is part of the backing for the device. Institutional assurances no doubt come in many forms, but for Cochrane Reviews, the most obvious of these are the various stages of peer review that distribute responsibility for the quality of a field's work products. A Cochrane Review requires prior approval of the constitution of the reviewer team and of the review protocol, for example, enlisting peer approval from the earliest stages of the work.

These resources are meant as strengtheners of the expert argument, but they are also, very often, a system of delegations (Jackson 2015a) in which responsibility for the validity of any one conclusion has been spread throughout a huge collective of participants. Jackson described delegation as assignment of a question or decision to some individual or group of individuals who can be trusted to exercise due diligence in the search for an answer—and delegation always involves accountability. Delegation occurs not only when society depends on expert communities, but also when expert communities depend on particular members to act on behalf of the whole.⁷ The individual performers of Cochrane Reviews take responsibility for faithful adherence to Cochrane procedures, but responsibility for the exhaustiveness of the search is delegated to the curators of databases; the responsibility for what is available to be retrieved has long since been delegated to funding agencies that set

⁷ Delegation is not limited to science-based arguments. As Jackson (2015a) pointed out, delegation is a general design feature that can be used in many situations, including most obviously the use of juries in trials at law.

research priorities and individual research teams that conduct primary research; the responsibility for establishing hierarchies of evidence is delegated to trusted working groups within Cochrane; the responsibility for approving a particular reviewing protocol is delegated to specific individual referees. No one in contemporary society, even experts, escapes the need to depend on others, as Willard (1990) pointed out. We argue that these delegations, which are undeniable features of contemporary argumentation, need some form of explicit consideration in both the analysis and the appraisal of arguments.

Other warranting devices will obviously differ in their design, and each one will have some distinctive set of assurances as backing. A more abstract representation of warranting devices appears in Fig. 3, suggesting that any such device may have backing that includes its own combination of material, procedural, and institutional assurances. The dotted line shows that the device is to be considered as the inference rule and all of its backing components; these are products of design that work together to warrant conclusions.

The most distinctive differences between warranting devices and familiar argumentation schemes are their field-specificity and their openness to redesign (Jackson 2015b). The primary purpose of a warranting device is to provide convincing evidence for a conclusion to people who understand the workings of the device and already have confidence in it. The way a community learns to have confidence in any kind of inference rule is through raising and responding to questions and objections. If something can be questioned on different bases in different communities, it has to earn its status in each such community, against each set of criteria. Cochrane Reviews belong to a well-defined context consisting of a primary readership composed of medical experts, a pre-existing literature, and other circumstantial features whose argumentative relevance is as yet unclear. The Cochrane device has developed iteratively from critique within the field, and it is still being elaborated to eliminate vulnerabilities in the conclusions it generates for any particular line of inquiry. Warranting devices, then, demand consideration of context, including not just the composition of the community within which they emerge but also the state of play within that community.

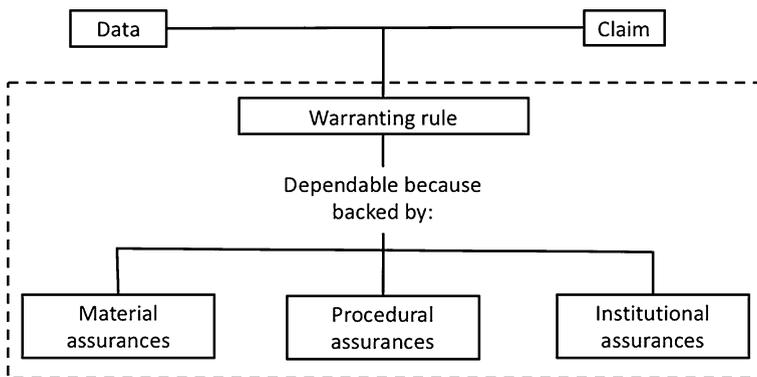


Fig. 3 Representing any warranting device in a Toulmin model

Contrary to what has often been assumed, though, field dependence does not mean that arguments warranted by a field's own devices can only be critiqued against the field's own standards. Empirically, this just is not the case. Cross-field critique can and does happen. The output of a device may become data for an appeal to expert opinion, and when that happens, the resulting argumentation can include authentic challenges to the validity of the device itself or to its use for a given purpose. Detailed empirical analysis of how non-experts engage with the devices of experts cannot be included here, but in the next section we offer some informal observations based on journalistic and public reception of Cochrane Reviews when their results appear in forms such as argument from expert opinion.

6 Device-Warranted Conclusions in the Hands of Non-experts

Any warranting device employed by an expert community will produce expert conclusions, and it is fully intended by Cochrane in particular that its own expert conclusions will serve as grounds for further conclusions drawn by health policy-makers, health care providers, and even individual citizens. In other words, the conclusions of Cochrane Reviews may become data for other arguments, including most obviously arguments from expert opinion. Walton (1997, pp. 19–20) points out the difference between evaluating the reasoning an expert has for an opinion and evaluating arguments from expert opinion that are presented to non-experts. Wagemans (2016a) formulates this difference in terms of first- and second-order predication: “the standpoint originally defended by the speaker—which consists of a subject and a predicate—is interpreted as the subject of the reconstructed standpoint.” An expert E has a standpoint, A, for which the grounds may be scientific. The fact that E holds that standpoint is a different kind of information, which may become a different kind of basis for belief by non-experts, leading to the following argumentation scheme (adapted from Walton et al. 2008, p. 310):

Major Premise: Source E is an expert in subject domain S containing proposition A.

Minor Premise: E asserts that proposition A is true.

Conclusion: A is true.

The Demicheli et al. review has been mentioned many times in connection with the anti-vaccination controversy (Jackson and Lambert 2016), often in contexts that closely resemble argument from expert opinion. For example, *The Guardian* included mention of the review in its “essential guide” to measles and MMR, as evidence that MMR vaccination does not cause autism:⁸

The evidence of no link between MMR and autism is now extremely strong. In February 2012, the Cochrane Collaboration—which compiles gold-standard

⁸ <http://www.theguardian.com/society/2013/apr/25/measles-mmr-the-essential-guide> captured Oct. 30, 2016, and saved as <https://perma.cc/FJ66-5CH7>. This appears in a series called “The Guardian’s essential guides: Everything you need to know to understand the biggest news stories.” In other words, this is not just a single news story but a well-edited “guide” that presumably is meant to be authoritative.

reviews of medical evidence—conducted a huge study into the safety of MMR. This mega-review brought together evidence from 54 difference [sic] scientific studies using a variety of methodologies and involving 14.7 million children from around the world. The study found “no association” between MMR and autism or a range of other conditions (asthma, leukaemia, hay fever, type 1 diabetes, gait disturbance, Crohn’s disease, demyelinating diseases, or bacterial or viral infections).

Here, A is the claim that “the evidence of no link between MMR and autism is now extremely strong.” The E identified in the passage is the Cochrane Collaboration and is asserted to be a compiler of “gold-standard reviews of medical evidence.” E’s opinion is paraphrased in the last sentence. If the reporting is accurate, the Cochrane Collaboration seems to have a considerable body of evidence to back the conclusion as stated by *The Guardian*.

In *The Guardian*’s full presentation, the appeal to the Cochrane Review is actually just one of several appeals to expert opinion, each of which is part of “the evidence of no link between MMR and autism.” Since each of these appeals to expert opinion might provide independent grounds for believing the claim, a complete examination of *The Guardian*’s argument would need to consider the soundness of each of these appeals. But here we focus only on the first (and superficially strongest) of the appeals to illustrate how an invented warranting device like Cochrane Review operates outside its own primary context.

When the conclusions of a Cochrane Review become grounds for an argument from expert opinion (as in the case of *The Guardian*), many if not all of the critical questions commonly associated with this scheme apply.⁹ Walton et al. (2008, p. 310) suggest the following list of critical questions for this scheme:

- Expertise question: How credible is E as an expert source?
- Field question: Is E an expert in the field that A is in?
- Opinion question: What did E assert that implies A?
- Trustworthiness question: Is E personally reliable as a source?
- Consistency question: Is A consistent with what other experts assert?
- Backup evidence question: Is E’s assertion based on evidence?

For any argument from expert opinion, Walton et al. (2008, p. 310) include a critical question known as the “opinion question”: What did the expert say that implies the proposition attributed to the expert? To answer this question requires going to the expert source, in this case, to the Cochrane Review itself. The review (Demicheli et al. 2012) did in fact look at 54 studies, but these 54 studies are divided among several different analyses involving a variety of adverse effects (some having to do with autism, some with other conditions). As already mentioned in Sect. 5, only 10 of the 54 studies looked at autism as a possible adverse reaction to vaccination; the other studies looked at other adverse reactions or only at efficacy. The studies relevant to autism are listed in Table 1 (together with a selection of other information provided in Table 9 of the review). Studies are grouped by design type;

⁹ Several of the questions would have to be asked quite differently to be meaningful in this context (the expertise, field, and consistency questions in particular).

Table 1 Assessments of evidence quality in 10 studies relevant to *The Guardian's* claim

Study by design type	Risk of bias	Generalizability
<i>Retrospective cohort designs</i>		
Madsen 2002	Moderate	High
Fombonne 2001	High	Low
Uchiyama 2007	High	Low
<i>Case-control designs</i>		
Smeeth 2004	Moderate	Medium
DeStefano 2004	Moderate	Medium
<i>Time-series designs</i>		
Mrozek-Budzyn 2010	Moderate	Medium
Fombonne 2006	High	Medium
<i>Self controlled case series</i>		
Honda 2005	Moderate	Medium
Makela 2002	Moderate	Medium
Taylor 1999	Moderate	Medium

Table entries are simplified data from Demicheli et al. (2012, Table 9)

none are RCTs (unremarkable since research ethics would prohibit randomly assigning children to receive or not receive vaccination). The reviewers' assessments of the quality of the individual studies (their risk of bias and the generalizability of their results) are shown in the second and third columns of the table, and it is by no means obvious that the reviewers themselves would agree that the 10 studies relevant to this particular claim provide "extremely strong" evidence. Reviewers classified all ten of the autism-related studies as containing either "high" risk of bias (meaning that important controls were missing or that serious design weaknesses were present) or "moderate/unknown" risk of bias (appearing simply as "moderate" both in the table findings of Demicheli et al. and in our Table 1). Studies were also scored for generalizability (or in the language of experimental design, for "external validity," commonly understood in terms of the match of a study population to a target population and/or the match of an experimental treatment to real-life circumstances the treatment is meant to represent). One of the ten studies was judged high in generalizability, two were judged low in generalizability, and the other seven were judged as having medium generalizability. These levels of generalizability across ten individual studies would likely be considered fairly persuasive for the ten taken together, if at least some of the studies to be generalized carried "low" risk of bias.

The Guardian, not the Cochrane Review itself, claims that the evidence on this point was strong. By contrast, the authors of the review were careful to point out the underlying weakness of the literature being reviewed; while they found no evidence of association between MMR and autism, they did not claim that the available research provides extremely strong evidence of an *absence* of association. According to the plain language summary (a required component of a Cochrane Review):

We could assess no significant association between MMR immunisation and the following conditions: autism, asthma, leukaemia, hay fever, type 1

diabetes, gait disturbance, Crohn’s disease, demyelinating diseases, or bacterial or viral infections. The methodological quality of many of the included studies made it difficult to generalise their results.

But scrutiny of arguments warranted directly or indirectly by Cochrane Review need not, and should not, end with critique based on lay interpretations of expert opinion. Any downstream argument derived from a warranting device will be vulnerable not only to generic challenges associated with (say) expert opinion, but also to challenges specific to the device itself.

A general class of such challenges might have to do with biases built into the device. A warranting device is always designed to answer some set of questions but not others, and it will usually assume those things that its expert users assume. To illustrate, a common notion within anti-vaccination discourse (characterized as conspiracy thinking by Oliver and Wood 2014) is that the institutions responsible for the production of the primary research have so strong an interest in mass immunization that they conceal or suppress evidence of serious risks. While no one expects scientists to engage with conspiracy theorists, it is certainly reasonable to ask what interests and assumptions shared by members of an expert community might make that community blind to certain evidence or deaf to certain arguments. There may also be reasonable questions to ask about the institutional assurances that back the device—whether, for example, anyone in the entire chain of delegations has really been motivated to search for a link between autism and MMR.

Another general class of such challenges might have to do more directly with what evidence the device is capable of “ingesting.” By design, a Cochrane Review ignores evidence that could (at least in principle) be relevant. This confers both strengths and limitations. A significant feature of the current design of the Cochrane Review is that it aggregates published (and sometimes unpublished) scientific research—presumably the best available scientific evidence. This is a strength. But it is also a potential limitation—an argumentative weakness even if not a scientific one, since on any given topic, there may be forms of evidence that are external to the scientific literature. Excluded are a very wide range of evidence types that can be supplied by ordinary people paying attention to their own health and their own reactions to treatments. This includes observation “pools” that are beginning to appear, whether as patient-to-patient sharing (Kazmer et al. 2014) or as aggregation of self-quantification activities (Fawcett 2015). In the vaccination controversy, the most notable such source is parents’ firsthand reports of their children’s reactions to vaccination; these are often quite credible to other parents, and may be weighed (in the public discussion) against the conclusions drawn from research reports in the scientific literature. One special source of data related to the vaccination controversy is the VAERS database—the Vaccination Adverse Events Reporting System that is used by the medical community to monitor vaccine safety, but that has also become an open information resource that activists have mined for support for their anti-vaccination views.¹⁰ These various forms of information are of greatly varying

¹⁰ It is worth noting that a scientific review *can* treat these reports as data and draw conclusions from them, as happened, for example, in a review commissioned by the Institute of Medicine in 2004 (released as *Immunization Safety Review: Vaccines and Autism*). In that review, VAERS reports were examined,

value in inferring causality, but they are part of what circulates in social media, competing for attention with press releases and other reporting of scientific news. Not considering them within a Cochrane Review may be a sensible choice within a community of experts, but it does not make the expert argument immune from questioning around the evidence that has been excluded. Limitations due to excluded evidence may affect many if not most warranting devices. Warranting devices of all kinds might be designed purposely to work with a specific kind of evidence but offer no capacity at all for working with other material, and whether this opens a conclusion to criticism may depend on circumstances like what other evidence is available on the topic.

An important point to notice is that critical questions about a particular device may come from any source, including from non-experts. An intelligent non-expert can pursue critical avenues connected to these device-specific issues. That is, at least some members of the general public may have the skills and motivation to engage in reasoning about the strength and credibility of the device. Not all of their questions and challenges will “draw blood” against the device, but when they do draw blood, that can give rise to further development of the device itself. Criticism, in other words, is a driver of future improvements and should be neither suppressed nor ignored.

7 The Field-Independent Status of Field-Dependent Devices

Before concluding we must return to the concern sometimes expressed about Toulmin’s theory, that if field dependence is acknowledged as a fact about argumentation, there may be no escape from some form of relativism. Objectionable to many on philosophical grounds, relativism also presents practical dilemmas: Expert communities might plausibly insist that their arguments are above criticism by non-experts, and in fact, they sometimes do. As we hope we have already shown, claims to immunity from critique actually avail nothing to an expert community. Although a warranting device may be applied in a completely uncontroversial way within an expert field, the device itself may attract all kinds of criticism either from within the expert field or from without. New problems may be noticed at any time, either by expert users of the device or by outsiders who become interested for any reason. Some knowledge of how the device operates may be needed to make an effective critique, but as we have already tried to show, interested non-experts can raise legitimate doubts about the conclusions said to be warranted by a device.

The fact that a device has earned the confidence of a group of experts is not generally sufficient to earn trust from other potential audiences. The testing ground for any new warranting device is argumentation itself, and any new question, regardless of source, is a new test to pass. The device must earn and continuously

Footnote 10 continued

along with published research, and effort was spent considering what inferences, if any, could be made from case reports. Our point here is of course not that unverified observations are good grounds for rebuttal of a Cochrane Review, but that Cochrane Reviews, by design, do not include all *possible* evidence on a question (as a consensus panel or a freeform debate might do).

maintain its status by withstanding critique, not only within its originating field, but also in each context to which it spreads.

Seeing warranting devices as encapsulations of how the expert community reasons, questions can be asked not only about the individual use of the device in one argument, but also about the assumptions the device encapsulates. This is an important shift of scale that involves questions that may need to be asked to correct an unsuspected bias in expert reasoning. Such questions can sometimes be formulated more easily by non-experts than by the experts themselves, by coming from a perspective with assumptions different from those shared within a field. For example, Jackson and Lambert (2016, pp. 548–549) described an incident during a National Academies workshop in which a motivated and well-informed member of the public was able to call out an unfounded assumption medical researchers had been making, in effect asserting that the medical research community was refusing to meet an obvious burden of proof, on the question of whether autism is, or is not, increasing in prevalence.

Accepting that reasoning may require field-specific standards for its evaluation may indeed seem to preclude critique from outside the field, and experts themselves often have the sense that outside critique is a form of interference. In one sense, this resistance to critique is justified: Non-experts lack the tacit knowledge shared within an expert community (see Collins and Evans 2007, for complete treatment of this issue), and their objections to expert argument can be badly misdirected. Direct public engagement with scientists can cross the boundary between legitimate questioning and illegitimate pressure (as pointed out by Lewandowsky and Bishop 2016). But at the same time, there is real danger in experts themselves coming to believe that their arguments are above all critique by non-experts. In health contexts where much is at stake, both experts and non-experts must fully explore the possible grounds for disagreement with conclusions drawn from experts’ devices, sometimes leading to material improvement in the devices themselves. This possibility of continuous improvement means that invention of new devices will always hold some promise of improvement in human reasoning as a whole. But the hardening of adherence to these devices presents a corresponding challenge: a tendency to dismiss without consideration new challenges that really merit a response.

Whether to embrace field-dependent reasoning as a positive contribution to human reasoning or to fear the power it gives to expert communities is an unsolved problem for argumentation theory, mirrored by the intractable practical problems that arise so frequently when expert reasoning and experts’ work products are drawn into public controversies. We have tried to emphasize in this case study that it is not just possible, but also necessary, that experts’ devices be critically examined by non-experts. Thus, one important goal in modeling warranting devices is to expose avenues for productive examination of the devices by non-experts, and another, equally important, is to support productive response by expert communities.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Aagaard, Thomas, Hans Lund, and Carsten Juhl. 2016. Optimizing literature search in systematic reviews: Are MEDLINE, EMBASE and CENTRAL enough for identifying effect studies within the area of musculoskeletal disorders? *BMC Medical Research Methodology* 16(1): 161–172. doi:10.1186/s12874-016-0264-6.
- Balshem, Howard, Mark Helfand, Holger J. Schünemann, Andrew D. Oxman, Regina Kunz, Jan Brozek, Gunn E. Vist, Yngve Falck-Ytter, Joerg Meerpohl, and Susan Norris. 2011. GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology* 64(4): 401–406. doi:10.1016/j.jclinepi.2010.07.015.
- Bermejo-Lugue, Lilian. 2006. Toulmin's model of argument and the question of relativism. In *Arguing on the Toulmin model*, ed. David Hitchcock and Bart Verheij, pp. 71–85. Dordrecht: Springer Netherlands. doi:10.1007/978-1-4020-4938-5_6.
- Bero, Lisa, and Drummond Rennie. 1995. The Cochrane Collaboration: Preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Journal of the American Medical Association* 274(24): 1935–1938. doi:10.1001/jama.1995.03530240045039.
- Boring, Edwin G. 1954. The nature and history of experimental control. *The American Journal of Psychology* 67(4): 573–589. doi:10.2307/1418483.
- Hill, Austin Bradford. 1965. The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine* 58(5): 295–300. doi:10.1177/0141076814562718.
- Chalmers, Iain, Murray Enkin, and Marc J.N.C. Keirse. 1989. *Effective care in pregnancy and childbirth: Pregnancy*. Oxford, UK: Oxford University Press.
- Chesñevar, Carlos, Sanjay Modgil, Iyad Rahwan, Chris Reed, Guillermo Simari, Matthew South, Gerard Vreeswijk, and Steven Willmott. 2006. Towards an Argument Interchange Format. *The Knowledge Engineering Review* 21(4): 293–316. doi:10.1017/S0269888906001044.
- Cochrane, Archie L. 1972. *Effectiveness and efficiency: Random reflections on health services*. London: Nuffield Provincial Hospitals Trust.
- Cochrane Crowd. n.d. <http://crowd.cochrane.org/index.html>. Accessed 15 July 2017.
- Cochrane Library. n.d. CENTRAL creation details. <http://www.cochranelibrary.com/help/central-creation-details.html>. Accessed 15 July 2017.
- Contaxis, Nicole. 2016. Grateful Med: Personal computing and user-friendly design. Circulating Now [blog of the U.S. National Library of Medicine]. <https://circulatingnow.nlm.nih.gov/2016/04/28/grateful-med-personal-computing-and-user-friendly-design/>. Accessed 15 July 2017.
- Collins, Harry M., and Robert Evans. 2007. *Rethinking expertise*. Chicago, IL: University of Chicago Press.
- Cooper, Harris M., and Robert Rosenthal. 1980. Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin* 87(3): 442–449. doi:10.1037/0033-2909.87.3.442.
- Demicheli, Vittorio, Alessandro Rivetti, Maria Grazia Debalini, and Carlo Di Pietrantonj. 2012. Vaccines for measles, mumps and rubella in children. *Cochrane Database of Systematic Reviews* 2: CD004407. doi:10.1002/14651858.CD004407.pub3.
- Dickersin, Kay, Eric Manheimer, Susan Wieland, Karen A. Robinson, Carol Lefebvre, and Steve McDonald. 2002. Development of the Cochrane Collaboration's CENTRAL Register of Controlled Clinical Trials. *Evaluation and the Health Professions* 25(1): 38–64. doi:10.1177/016327870202500104.
- Dickersin, Kay, Roberta Scherer, and Carol Lefebvre. 1994. Identifying relevant studies for systematic reviews. *British Medical Journal* 309(6964): 1286–1291.
- Fawcett, Tom. 2015. Mining the quantified self: Personal knowledge discovery as a challenge for data science. *Big Data* 3(4): 249–266. doi:10.1089/big.2015.0049.
- Freeman, James B. 2005. Systematizing Toulmin's warrants: An epistemic approach. *Argumentation* 19(3): 331–346. doi:10.1007/s10503-005-4420-0.
- Garritty, Chantelle, Adrienne Stevens, Gerald Gartlehner, Valerie King, Chris Kamel, and Cochrane Rapid Reviews Methods Group. 2016. Cochrane Rapid Reviews Methods Group to play a leading role in guiding the production of informed high-quality, timely research evidence syntheses. *Systematic Reviews* 5(1): 184–188. doi:10.1186/s13643-016-0360-z.
- Glass, Gene V. 1976. Primary, secondary, and meta-analysis of research. *Educational Researcher* 5(10): 3–8. doi:10.3102/0013189x005010003.

- Grant, Maria J., and Andrew Booth. 2009. A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal* 26(2): 91–108. doi:[10.1111/j.1471-1842.2009.00848.x](https://doi.org/10.1111/j.1471-1842.2009.00848.x).
- Guyatt, Gordon H., Andrew D. Oxman, Holger J. Schünemann, Peter Tugwell, and Andre Knottnerus. 2011. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *Journal of Clinical Epidemiology* 64(4): 380–382. doi:[10.1016/j.jclinepi.2010.09.011](https://doi.org/10.1016/j.jclinepi.2010.09.011).
- Harlan, William D. 1993. An evidence based health care system: The case for clinical trials registries. Report on a National Institutes of Health Technology Assessment Workshop, Bethesda, Maryland, December 6–7, 1993. <https://consensus.nih.gov/1993/1993EvidenceBasedTrialRegistriesta013html.htm>. Accessed 15 July 2017.
- Hedges, Larry V. 1986. Issues in meta-analysis. *Review of Research in Education* 13: 353–398. doi:[10.3102/0091732X013001353](https://doi.org/10.3102/0091732X013001353).
- Higgins, Julian P. T., and Sally Green. 2011. Cochrane handbook for systematic reviews of interventions: The Cochrane Collaboration. <http://handbook.cochrane.org>. Accessed 14 Jan 2017.
- Hitchcock, David. 2003. Toulmin’s warrants. In *Anyone who has a view: Theoretical contributions to the study of argumentation*, ed. Frans H. van Eemeren, 69–82. Dordrecht: Kluwer Academic. doi:[10.1007/978-94-007-1078-8_6](https://doi.org/10.1007/978-94-007-1078-8_6).
- Hitchcock, David. 2005. Good reasoning on the Toulmin model. *Argumentation* 19(3): 373–391. doi:[10.1007/s10503-005-4422-y](https://doi.org/10.1007/s10503-005-4422-y).
- Institute of Medicine. 2004. *Immunization safety review: Vaccines and autism*. Washington, DC: The National Academies Press.
- Jackson, Sally. 2015a. Deference, distrust, and delegation: Three design hypotheses. In *Reflections on theoretical issues in argumentation theory*, 227–243. Springer. doi:[10.1007/978-3-319-21103-9_17](https://doi.org/10.1007/978-3-319-21103-9_17).
- Jackson, Sally. 2015b. Design thinking in argumentation theory and practice. *Argumentation* 29(3): 243–263. doi:[10.1007/s10503-015-9353-7](https://doi.org/10.1007/s10503-015-9353-7).
- Jackson, Sally, and Natalie Lambert. 2016. A computational study of the vaccination controversy. In *Argumentation and Reasoned Action: Proceedings of the First European Conference on Argumentation, Lisbon, 9-12 June 2015, Vol.II.*, ed. Dima Mohammed & Marcin Lewinski, 539–552. College Publications (Studies in Logic and Argumentation).
- Kazmer, Michelle M., Liza A. Mia, Juliann Cortese Lustria, Gary Burnett, Ji-Hyun Kim, Jinxuan Ma, and Jeana Frost. 2014. Distributed knowledge in an online patient support community: Authority and discovery. *Journal of the Association for Information Science and Technology* 65(7): 1319–1334. doi:[10.1002/asi.23064](https://doi.org/10.1002/asi.23064).
- Keith, William M., and David E. Beard. 2008. Toulmin’s rhetorical logic: What’s the warrant for warrants? *Philosophy and Rhetoric* 41(1): 22–50. doi:[10.1353/par.2008.0003](https://doi.org/10.1353/par.2008.0003).
- Khangura, Sara, Kristin Konnyu, Rob Cushman, Jeremy Grimshaw, and David Moher. 2012. Evidence summaries: The evolution of a rapid review approach. *Systematic Reviews* 1: 10–18. doi:[10.1186/2046-4053-1-10](https://doi.org/10.1186/2046-4053-1-10).
- Lefebvre, Carol, L. Julie Glanville, Susan Wieland, Bernadette Coles, and Alison L. Weightman. 2013. Methodological developments in searching for studies for systematic reviews: Past, present and future? *Systematic Reviews* 2: 78–86. doi:[10.1186/2046-4053-2-78](https://doi.org/10.1186/2046-4053-2-78).
- Lefebvre, Carol, Eric Manheimer, and Julie Glanville. 2011. Searching for studies. In Higgins, Julian P. T., and Sally Green. 2011. *Cochrane handbook for systematic reviews of interventions: The Cochrane Collaboration*. <http://handbook.cochrane.org>. Accessed 14 Jan 2017.
- Lewandowsky, Stephan, and Dorothy Bishop. 2016. Research integrity: Don’t let transparency damage science. *Nature* 529(7587): 459–461. doi:[10.1038/529459a](https://doi.org/10.1038/529459a).
- Medical Research Council. 1948. Streptomycin treatment of pulmonary tuberculosis. *British Medical Journal* 2(4582): 769–782. doi:[10.1136/bmj.2.4582.769](https://doi.org/10.1136/bmj.2.4582.769).
- Mercier, Hugo. 2011. When experts argue: Explaining the best and the worst of argumentation. *Argumentation* 25(3): 313–327. doi:[10.1007/s10503-011-9222-y](https://doi.org/10.1007/s10503-011-9222-y).
- Mercier, Hugo, and Dan Sperber. 2011. Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences* 34(2): 57–74. doi:[10.1017/S0140525X10000968](https://doi.org/10.1017/S0140525X10000968).
- Moher, David, Lesley Stewart, and Paul Shekelle. 2015. All in the family: Systematic reviews, rapid reviews, scoping reviews, realist reviews, and more. *Systematic Reviews* 4: 183–184. doi:[10.1186/s13643-015-0163-7](https://doi.org/10.1186/s13643-015-0163-7).
- Noel-Storr, Anna, Gordon Dooley, Julie Glanville, and Ruth Foxlee. 2015. *The Embase project 2: Crowdsourcing citation screening*. Vienna, Austria: Cochrane Colloquium. Abstract at <https://abstracts.cochrane.org/2015-vienna/embase-project-2-crowdsourcing-citation-screening> Slides from

- https://www.researchgate.net/profile/Anna_Noel-Storr/project/The-Embase-project/attachment/572c56de08aea7adff2ed046/AS:358664184582145@1462523613962/download/FINAL_Embase_project_long_oral_2015_v1.pptx. Accessed 15 July 2017.
- Office of Technology Assessment (US Congress). 1982. MEDLARS and health information policy. Washington, DC: US Government Printing Office. <http://resource.nlm.nih.gov/101021663>.
- O'Keefe, Daniel J. 1999. Variability of persuasive message effects: Meta-analytic evidence and implications. *Document Design* 1(2): 87–97. doi:10.1075/dd.1.2.02oke.
- Oliver, J.Eric, and Thomas Wood. 2014. Medical conspiracy theories and health behaviors in the United States. *JAMA Internal Medicine* 174(5): 817–818. doi:10.1001/jamainternmed.2014.190.
- Pinch, Trevor J., and Wiebe E. Bijker. 1984. The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science* 14(3): 399–441. doi:10.1177/030631284014003004.
- Polisena, Julie, Chantelle Garritty, Chris Kamel, Adrienne Stevens, and Ahmed M. Abou-Setta. 2015. Rapid review programs to support health care and policy decision making: A descriptive analysis of processes and methods. *Systematic Reviews* 4: 26–32. doi:10.1186/s13643-015-0022-6.
- Rosenthal, Robert. 1984. *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.
- Synnot, Anneliese, Russell L. Gruen, David Menon, Ewout W. Steyerberg, Andras Buki, Wilco C. Peul, Julian H. Elliott, and Andrew Maas. 2016. A new approach to evidence synthesis in traumatic brain injury: A living systematic review. *Journal of Neurotrauma*. doi:10.1089/neu.2015.4124.
- Toulmin, Stephen E. 1950. *An examination of the place of reason in ethics*. Cambridge: Cambridge University Press.
- Toulmin, Stephen E. 1958. *The uses of argument*. Cambridge: Cambridge University Press.
- Toulmin, Stephen E., Richard D. Rieke, and Allan Janik. 1984. *An introduction to reasoning*, 2nd ed. New York: Macmillan.
- Underhill, Kristen, Paul Montgomery, and Don Operario. 2007. Sexual abstinence only programmes to prevent HIV infection in high income countries: Systematic review. *BMJ (British Medical Journal)* 335: 248–252. doi:10.1136/bmj.39245.446586.BE.
- U.S. National Library of Medicine. MEDLINE FactSheet <https://www.nlm.nih.gov/pubs/factsheets/medline.html>. Accessed 15 July 2017.
- Wagemans, Jean H.M. 2014. Argumentation from expert opinion in the 2011 US debt ceiling debate. In *Disturbing argument: Selected works from the 18th NCA/AFA Alta Conference on Argumentation*, ed. Catherine Palczewski, 49–56. Abingdon: Taylor & Francis.
- Wagemans, Jean H.M. 2016a. Argumentation patterns for justifying scientific explanations. *Argumentation* 30(1): 97–108. doi:10.1007/s10503-015-9374-2.
- Wagemans, Jean H.M. 2016b. Constructing a periodic table of arguments. In *OSSA Conference Proceedings* 11, 106, 1–12, <http://scholar.uwindsor.ca/ossaarchive/OSSA11/papersandcommentaries/106/>.
- Walton, Douglas N. 1997. *Appeal to expert opinion: Arguments from authority*. University Park, PA: Pennsylvania State University Press.
- Walton, Douglas, Chris Reed, and Fabrizio Macagno. 2008. *Argumentation schemes*. Cambridge: Cambridge University Press.
- Willard, Charles Arthur. 1990. Authority. *Informal Logic* 12(1): 11–22.
- Young, Diony. 1990. Review of *Effective Care in Pregnancy and Childbirth* Edited by Iain Chalmers, Murray Enkin, and Marc J.N.C. Keirse. *Birth* 17(1):55–62. doi:10.1111/j.1523-536X.1990.tb00014.x.
- Zarefsky, David. 1982. Persistent questions in the theory of argument fields. *Journal of the American Forensic Association* 18(4): 191–203. (Reprinted in *Rhetorical perspectives on argumentation: Selected essays*. 2014. New York: Springer.).
- Zwahlen, Marcel, Andrew Renehan, and Matthias Egger. 2008. Meta-analysis in medical research: potentials and limitations. *Urologic Oncology: Seminars and Original Investigations* 26(3): 320–329. doi:10.1016/j.urolonc.2006.12.001.