



OPEN ACCESS

Different knowledge, different styles of reasoning: a challenge for guideline development

Sietse Wieringa,^{1,2} Dunja Dreesens,^{3,4} Frode Forland,⁵ Carel Hulshof,⁶ Sue Lukersmith,⁷ Fergus Macbeth,⁸ Beth Shaw,⁹ Arlène van Vliet,¹⁰ Teun Zuiderent-Jerak,¹¹ on behalf of the AID Knowledge Working Group of the Guidelines International Network

10.1136/bmjebm-2017-110844

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjebm-2017-110844>).

For numbered affiliations see end of article.

Correspondence to:
Dr Sietse Wieringa,
Department of Continuing
Education, University of
Oxford, Oxford OX1 2JA, UK;
sietse.wieringa@kellogg.ox.ac.uk

Introduction: the challenge of knowledge inclusion in guidelines

Evidence-based guidelines whether national, regional or developed by specialty groups, must search for, and explicitly consider, evidence from sources other than conventional clinical trials and their quantitative data. This need for appraising and including knowledge from a wide variety of sources in guideline development is well recognised.¹⁻³

Although evidence on statistical association—usually from randomised controlled trials (RCTs)—is commonly thought to be the dominant type of knowledge appraised and included, guideline developers frequently use a range of other types of knowledge including the views and experiences of those using and providing health services, understanding of how interventions work (eg, from logic models or realist evaluations), and other information, such as aetiology and the context of care (online supplementary text box 1).

These different types of knowledge are used and needed in many situations, for example, when evidence from RCTs is not available, impossible to obtain, contradictory or inappropriate. They can also be used in conjunction with knowledge from RCTs to provide context, to assess relevance and to understand bias. Furthermore, explicit (written or spoken) knowledge and the more intricate forms of knowledge like experiential and contextual knowledge can help guideline makers to take an approach consistent with the intentions of early evidence-based medicine (EBM) proponents: namely, that best evidence is not restricted to evidence from RCTs and meta-analyses alone.⁴

However, how to properly appraise (judge) and include (integrate) different kinds of knowledge remains unclear. Agreed methods are not yet available or are in the early stages of development and the need for and use of different kinds of knowledge is not always explicitly acknowledged, which affects the use of guidelines in practice.^{5,6} International and cultural differences in guideline production practices may further impede developments in appraising and including a broader range of types of knowledge (online supplementary text box 2).

In this paper, we discuss four specific aspects of guideline development to highlight the main challenges identified by the AID Knowledge Working Group through discussions and workshops with guideline developers and users (online supplementary text box 3):

1. the purpose of guideline development;
2. the problem of induction;
3. the dominance of frequency based reasoning;
4. the challenge of integrating different sources of knowledge.

In order to do this, we refer to some philosophical concepts around knowledge creation.

The purpose of guideline development

The efforts of the pioneers of the EBM movement were primarily in response to the discovery of the variation problem in population studies. Reducing variation of the care provided at a population level was considered to be an important way to achieve improved quality for individual patients.⁷ Hence, epidemiology, the science of studying *populations*, gained prominence in guidelines, the aims of which are to support decisions for *individual* patients. Classic epidemiology became clinical epidemiology when introduced to the bedside and the dominance of RCTs as the gold standard for intervention studies to assess causal relation between interventions and effect followed in this construct of epidemiology as used in EBM. The underlying—yet little explored—assumption is that guidelines based on population studies provide the best advice to inform clinical decisions for individual patients or situations.

However, reducing variation is not the only reason for developing guidelines; they are developed for several reasons, of which the most important one is to improve the quality of care. In order to meet the range of needs, guidelines may need different approaches, such as summarising large quantities of knowledge for practising healthcare professionals, serving as an intermediate product for other tools or applications (such as clinical decision support software) or providing implementation guidance. Although not primarily developed for this purpose, guidelines can also serve as tools to legally shield both patients and professionals, to help governments and health

Check for updates

To cite: Wieringa S, Dreesens D, Forland F, et al. *BMJ Evidence-Based Medicine* 2018;**23**:87–91.

insurers allocate scarce resources and to act as governance frameworks for practitioners and governments.

There is also the role of guideline development as a discipline in itself; along with its associated practices and institutions, it provides employment and intellectual interest for many.

There has been surprisingly little research into the purposes of guideline development. One mixed-method study found the purposes of guidelines were: defining norms, summarising evidence, formalising current consensus and/or describing current practices in a handbook-type format.⁵

Making the purposes of guidelines more explicit may help determine how different types of knowledge could and should be used. For instance, if the aim is to describe current good practice (eg, how services are organised to deliver care), this may be better achieved by drawing on qualitative or mixed-method evaluative research rather than RCTs. If the aim is to assess the effectiveness of a specific treatment or approach, evidence from RCTs or high-quality prospective cohort studies would usually be the primary source of knowledge, with qualitative or mixed-method studies serving to help understand the local context of implementation.

It is important to note, however, that for most guideline developers, the primary purpose remains that of supporting decision making in the clinical encounter. This leads us to the next fundamental aspect of guideline development.

The problem of induction

How do different types of knowledge in guidelines development help to make clinical decisions? Some basic concepts from the philosophy of science may help to understand the problem.

Inference, the problem of induction and evasions

In logic, to *infer* means to conclude from evidence using reasoning.⁸ In everyday healthcare practice, care professionals and patients reason to reach conclusions about what has happened, to make predictions about what will happen and to decide what to do next. Because of uncertainty in medicine, we usually deal with a specific type of inference, called *induction*, where the conclusions of our reasoning are not always right even when based on true premises. In philosophy, there is a concern whether this is actually possible, called the *problem of induction*,⁸ as introduced by Hume in 1739.⁹ At its simplest, this means we cannot predict the future with certainty. Although this seems reasonable, we are in fact able to predict the future quite accurately on many occasions in clinical practice. How is this possible? Philosopher of science Ian Hacking⁸ argues that we never solve the problem of induction, but only *evade it* by applying different kinds of reasoning to reduce uncertainty and increase our chance of reaching the best possible outcome.

The dominance of frequency-based reasoning

The evasion most dominantly used in guideline development is frequency-type reasoning in the form of systematic reviews, RCTs and observational studies.⁵ This evades the problem of induction by recognising that ‘although we can’t predict the future for the individual case, we can be “usually” right (eg, 95% of the time)⁸ as long as events or cases are frequent enough.

Frequency-based reasoning relies on basic assumptions that have some drawbacks. First, this line of reasoning assumes that reality is dice like and that we—eg, scientists guideline developers and healthcare professionals—are rolling the same dice (online supplementary text box 4). Frequency-type reasoning presupposes adequate framing and defining of what is similar and what is not, which is always based on judgement and choice.

Second, frequency-based reasoning aims to find simple causal correlations, independent of context. The question is whether these simple correlations hold true in real life. Different understandings of causality exist that could help us address this drawback.¹⁰ For instance, a network of complex causal relationships may be more realistic. This drawback is described as the *efficacy paradox*, where the different interference from non-specific effects (different from those controlled for between groups in a trial), measurement artefacts (that mimicked therapeutic effects in the trial) and regression patterns (such as the self-limiting nature of a disease) in real life can outweigh the specific effect found in a trial. This paradox may become especially apparent when inferring in the context of multimorbidity.¹¹

Finally, and most importantly, although frequency-based reasoning works well for frequent events (large groups, many data points and long periods of time), such reasoning faces fundamental limitations when inferring in the single-case scenario: a single patient, a rare disease, a system intervention and a one-off event. This can be particularly challenging when recommendations based on frequency-type evidence alone are deployed to help decision making for individual patients or unique situations, such as a public health response to a disease outbreak.²

Given these drawbacks, it is worth noting that other types of reasoning to evade the epistemological problem of induction exist.

In table 1, several alternative ways of reasoning are listed. They are mainly used in areas where frequency-based reasoning is particularly problematic, for instance in guidelines focusing on complex interventions, public and occupational health, rehabilitation, and social care and welfare.^{11–13}

These different types of reasoning try to help make valid inferences for the single-case scenario, when there is no frequency of events. Many of these are already recognised and stated by Bradford-Hill¹⁴ in his criteria for causation, but some are newer, such as Annemarie Mol’s logic of care,¹⁵ where a practitioner will try something, wait and see and let unfolding events guide the next step. Using this type of reasoning, the problem of induction is solved through ‘tinkering’, making incremental changes to improve a situation.

Guidelines can and do support these kinds of evasions by including different types of knowledge. For instance, providing laboratory information about aetiology helps to make an inference based on mechanistic reasoning.¹⁶ A description of cases of harm can offer an inference based on the precautionary principle.¹⁷ Rethinking how inferences are made in practice may shift the dominance of frequency-based reasoning and its reliance on a restrictive type of knowledge to a broader spectrum of knowledge being used to support different reasoning approaches. The need for using different type of knowledge is shown by a large Dutch analysis showing that knowledge from RCTs far outweighed other knowledge types used, irrespective of the question at hand, thus ignoring important and relevant knowledge from other sources.^{5,6}

The challenge of integration

Making a recommendation for a specific healthcare problem in a specific healthcare system requires the assessment of knowledge not just on its own merits, but importantly its integration with other knowledge. Indeed, EBM is defined as integrating the best evidence with clinical expertise and patient preference.⁴

However, in the context of medicine, and even more so in that of guideline production, integration of different types of knowledge remains underexplored and undertheorised.

Some areas of evidence synthesis have addressed integration. For example, statistical techniques such as meta-analysis can be

Table 1 Alternative types of reasoning to evade the problem of induction

Type of reasoning (with examples of key scholars)	Shorthand description	Explanation
Bayesian evasion (Bayes, Hacking)	Learning from experience	This type of inductive inference agrees with Hume that we cannot predict the future perfectly, but that we can learn from our experiences reasonably well. This allows us to do more and better predictions. This type of reasoning can update current beliefs with information from frequent events (informing prior probabilities and likelihood ratios). However, because we can learn from a single event too, this approach is suited for the individual case scenario. ⁸
Abduction (Peirce)	Reasoning to the best explanation	Abduction makes inferences by updating beliefs leading to the best explanation. ²⁸ Where Bayesian evasion takes prior probabilities as a given (at least as some argue), which may not be the case, abduction does not. It introduces the consideration of theory and mechanism in the act of inferring. ²⁹
Mechanistic/deterministic reasoning	How things appear to work	This type of reasoning makes an inference based on a mechanism. Illari <i>et al</i> ¹⁶ define a mechanism as consisting 'of entities and activities organised in such a way that they are responsible for the phenomenon'.
Falsification (Popper)	Trial and error	Popper ³⁰ agreed with Hume: we cannot say anything about the future, there are only theories that cannot even be proven. At best, we can only prove that they are wrong (falsifiable). This 'anti-inductivist' reasoning suggests to continue using a certain theory or practice and adjust if they fail.
Precautionary principle	In case of uncertainty about the future prevent harm	The precautionary principle, often used in environmental decision making and occupational health, favours to take preventive action in the face of uncertainty when making an inference. It puts 'the burden of proof to the proponents of an activity; exploring a wide range of alternatives to possibly harmful actions; and increasing public participation in decision making'. ^{17 31}
Means-to-ends reasoning	Find ways to reach a goal	This type of reasoning asks the question what ways are there to reach a certain wanted outcome and which of those ways would be the more efficient? Often used in clinical consultations to make sure that something happens whatever the circumstances. The inference remains uncertain but less so by using multiple means that will lead to the same outcome. ³²
Logic of care (Mol)	Taking care while the uncertain future unfolds	In <i>The Logic of Care</i> Annemarie Mol ¹⁵ suggests that healthcare is more like a 'practice' than it is about making choices. This approach puts emphasis on the importance of taking good care for the patient and the prevention of neglect. Inferring is a process that unfolds over time, while addressing many factors on the way.
Non-analytical reasoning (Gigerenzer, Stolper)	Using intuition	Non-analytical reasoning such as heuristics and gut feelings (combination of heuristics and emotions ³³) used to make inferences. These types of reasoning are considered fast, intuitive and automatic thought processes. Gigerenzer showed that non-analytical reasoning can in certain environments outperform analytical reasoning in psychological, biological, sociological and economic inference tasks. ³⁴

Types of reasoning are not exclusive and may overlap.

used to combine data from different studies, and another range of techniques can be used to synthesise qualitative data.

In guideline development, most of the activities and tools to support high-quality evidence synthesis such as risk of bias assessment and quality assessment (such as GRADE) tend to focus primarily on frequency-based reasoning and knowledge. For the assessment of quality of qualitative evidence, there are limited but relevant initiatives for guideline development in progress, for example, the recently published Grading of Recommendations Assessment, Development, and Evaluation – Confidence in the Evidence from Reviews of Qualitative research (GRADE-CER-Qual)¹⁸ guidance. However, many of these efforts try to achieve integration by synthesizing studies that share the same questions and design (eg a set of qualitative or, more narrowly, ethnographic studies),¹⁹ at times appraising¹⁸ all such knowledge again in frequentist terms, like with some qualitative evidence synthesis methods²⁰ that 'emphasize frequencies of the qualitative data they present...undermin[ing] the uniqueness of the qualitative knowledge they proclaim by focusing on frequency and the general patterns'.²¹

The main issue is that these tools, activities and initiatives aim to integrate *similar* knowledge, such as data from the same study designs, the same populations or the same outcomes. How different kinds of knowledge are valued, appraised and weighed

in relation to each other, for example, regarding effectiveness, efficiency or ethical concerns, is not clearly articulated.

Nonetheless, guideline developers do recognise that other types of knowledge are often used and somehow integrated in practice, particularly when discussing the evidence and formulating recommendations, often called 'judgement' or 'considered judgement'.^{22 23} This is the traditionally less clearly described or analysed black box part of the process that new initiatives try to shed a light on, such as NICE's structured tables linking evidence to recommendations and the GRADE Evidence to Decision frameworks.²²

They appear promising yet challenges remain. First, findings from ethnography question whether structured frameworks really influence or reflect guideline development processes.^{24 25} In an ethnographic study of guideline development meetings, Moreira showed that guideline developers formulate guidance by combining different '*repertoires of evaluation*, organised around four different epistemic criteria: robustness, usability, acceptability and adequacy'.²⁴ Importantly, such criteria are deployed at each stage of evidence appraisal: usability, acceptability and adequacy are integral to evidence assessment, rather than being easily categorised as either 'judgements' or 'additional considerations' as current evidence to decision frameworks suggest.²² Acknowledging the importance of these epistemic skills in evidence

appraisal become much more important when it is understood that recommendations nearly always draw on different types of knowledge.

Second, bringing knowledge together is not just a process of integrating, triangulation and finding a single answer. Knowledge from many sources is often conflicting, and indeed the exploration of opposing ideas is often very important. In social sciences, methods for evidence synthesis of other kinds of data have been developed and assessed in research²⁶ and in practice guideline development,¹³ but these have not yet been adopted routinely in healthcare guideline development. A process of integration is not just a technical, simple mechanistic process. Guideline development is a human, social process involving relevant stakeholders in discussion, debate and judgement. Therefore, the guideline development processes also relies on a balanced and representative guideline committee that functions well.²⁷

Finally, integrating many types of knowledge is not a process in which anything goes. Some integration processes are likely to be better than others. Guideline development needs to be transparent and consistent so that reality, be it physical or social, can limit the inferences and recommendations made. We need a range of integration approaches depending on our understanding of what is true and real; for example, integration of different knowledge could be based on combinations of coherence (what fits best in a network of other theories), on consensus (what people agree on) and/or on correspondence (what links best to what is believed to be real).

Given the current state of evidence to decision frameworks, there is still little guidance on how to robustly and consistently combine knowledge of different types without using the frequentist understanding of knowledge. A broader discussion within the guideline community is needed about the frameworks used to integrate and include different kinds of knowledge. Considering theories from epistemology and findings from ethnography (online supplementary text box 5) could be instrumental to deepen our understanding of how other types of knowledge can be synthesised and integrated in guideline production.

Conclusion

The development of guideline recommendations is an interactive human process that requires a range of knowledge and experience including, but not exclusively, knowledge from frequency-based research, such as clinical trials. As in the clinical encounter, appraising and including different types of knowledge in guideline development should be used to make better inferences to guide decisions, but in practice, arguments are used to exclude some kinds of knowledge for a range of reasons, including concerns about introducing bias in frequentist reasoning.

In this paper, we present important epistemological reasons to appraise and include a (wide) variety of different types of knowledge to highlight important aspects of guideline development that await further exploration and practical suggestions.

We acknowledge that appraising and including knowledge from a different variety of sources is likely to be complex and ongoing. Discussions about purpose, reasoning and integration in guideline development will continue. A simple set of tools or methodological quick fixes are unlikely to suffice, and developing criteria for appraising and integrating different knowledge will remain a challenge. However, we believe that much can be done to help guideline developers improve this—now often implicit⁵ practice that is central to their work. Capacity-building workshops that confront implicit forms of reasoning are one example. AID Knowledge runs such workshops annually at G-I-N conferences.

They help to strengthen ties between guideline developers who are concerned about the increasingly rigid methodological constraints on guideline methods at the expense of fostering epistemic sensibilities. It is important for guideline developers to feel they are part of a community of practice that encourages epistemic skill development, rather than a hierarchical community where superior guideline methods are defined by a small group of experts. This will help to keep guideline development innovative and diverse.

Acknowledging that dominant frequentist methods are excellent for some questions but do not fit all knowledge needs is the first step to implementing different kind of reasoning in guideline development. How to address the diversity in methods for different kinds of questions should be among the top guideline research priorities.

Author affiliations

¹Department of Health Sciences, University of Oslo, Oslo, Norway

²Department of Continuing Education, University of Oxford, Oxford, UK

³School CAPHRI, department of General Practice, Maastricht University, Maastricht, The Netherlands

⁴Knowledge Institute of Medical Specialists, Utrecht, The Netherlands

⁵Division for Infectious Diseases and Environmental Health, Norwegian Institute of Public Health, Oslo, Norway

⁶Coronel Institute of Occupational Health, Amsterdam Medical Center, Amsterdam, The Netherlands

⁷Research School of Population Health, Australian National University, Australia

⁸Centre for Trials Research, Cardiff University, Cardiff, UK

⁹National Institute for Health and Care Excellence, London, UK

¹⁰Leiden University Medical Center/Dutch Working Party on Infection Prevention, Leiden, The Netherlands

¹¹Department of Thematic Studies - Technology and social change, Linköping University, Linköping, Sweden

Collaborators Stephanie Chang, Pwee Keng Ho, Sonja Kersten, Miranda Langedam, Peter O'Neill, Sarah Richards, Rodrigo Pardo Turriago, Sue Phillips.

Contributors All authors substantially contributed to the conception or design of the work, or the acquisition, analysis or interpretation of data; drafted the work or revising it critically for important intellectual content; gave final approval of the version published; agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding SW received funding from the European Union Seventh Framework Programme (FP7 PEOPLE 2013 COFUND), Grant/Award Number: 609020. TZJ received funding from a fellowship in the Future Research Leaders program of Linköping University. The OA fee was paid for by the Guidelines International Network.

Competing interests SW was a guidelines update standing committee member for NICE from 2014 until 2018. FF was a Board member of Guidelines International Network from 2009 - 2013. He led the Norwegian Guideline Secretariat from 2001 - 2006 and has been a member of several Guideline Working Groups nationally and internationally. CH is guideline co-ordinator of the Netherlands Society of Occupational Medicine. SL worked as a guideline methodologist and developer in Australia. FM worked with NICE on clinical guidelines since 2000 and was Director of the Centre for Clinical Practice (responsible for guidelines) at NICE from 2008-2011. BS worked for NICE from 2008 until 2017, advising on the methods of

guideline development. She is also the chair of the GIN AID Knowledge Working Group. TZJ was founding co-chair of the AID Knowledge Working Group. He has studied guideline development for about a decade, through multiple externally funded studies.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

References

- Burgers JS, van Everdingen JJ. Beyond the evidence in clinical guidelines. *Lancet* 2004;364:392–3.
- Zuiderent-Jerak T, Forland F, Macbeth F. Guidelines should reflect all knowledge, not just clinical trials. *BMJ* 2012;345:e6702.
- Salvador-Carulla L, Lukersmith S, Sullivan W. From the EBM pyramid to the Greek temple: a new conceptual approach to Guidelines as implementation tools in mental health. *Epidemiol Psychiatr Sci* 2017;26:105–14.
- Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71–2.
- Zuiderent-Jerak T, Jerak-Zuiderent S, Bovenkamp H, et al. *Variatie in Richtlijnen; wat is het probleem?* Instituut Beleid en Management Gezondheidszorg, Erasmus Universiteit Rotterdam, 2011.
- van de Bovenkamp HM, Zuiderent-Jerak T. An empirical study of patient participation in guideline development: exploring the potential for articulating patient knowledge in evidence-based epistemic settings. *Health Expect* 2015;18:942–55.
- Wennberg JE. Dealing with medical practice variations: a proposal for action. *Health Aff* 1984;3:6–32.
- Hacking I. *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge university press, 2001.
- Hume D. *A treatise of human nature*. London : Thomas and Joseph Allman, 1817
- Kerry R, Eriksen TE, Lie SA, et al. Causation and evidence-based practice: an ontological review. *J Eval Clin Pract* 2012;18:1006–12.
- Walach H, Falkenberg T, Fønnebo V, et al. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol* 2006;6:29.
- Harder T, Abu Sin M, Bosch-Capblanch X, et al. Towards a framework for evaluating and grading evidence in public health. *Health Policy* 2015;119:732–6.
- Beauchamp S, Drapeau M, Dionne C. The development of practice guidelines in the social and human sciences. *Canadian Psychology/Psychologie canadienne* 2015;56:357–67.
- Bradford-Hill A. The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine* 1965:295–300.
- Mol A. *The Logic of Care: Health and the Problem of Patient Choice*. Routledge, 2008.
- Illari PM, Williamson J. What is a mechanism? Thinking about mechanisms across the sciences. *Eur J Philos Sci* 2012;2:119–35.
- van Dijk FJ, Verbeek JH, Hoving JL, et al. A knowledge infrastructure for occupational safety and health. *J Occup Environ Med* 2010;52:1262–8.
- Lewin S, Glenton C, Munthe-Kaas H, et al. Using Qualitative Evidence in Decision Making for Health and Social Interventions: An Approach to Assess Confidence in Findings from Qualitative Evidence Syntheses (GRADE-CERQual). *PLoS Med* 2015;12:e1001895–18.
- Lewin S, Booth A, Glenton C, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. *Implement Sci* 2018;13:2–10.
- Ring N, Ritchie K, Mandava L, et al. *A guide to synthesising qualitative research for researchers undertaking health technology assessments and systematic reviews*: NHS Quality Improvement Scotland, 2011. 1-84404-917-5.
- Juritzen TI, Soberg HL, Røe C, et al. The One or the Many: Quantified Subjectivity and Aggregated Uniqueness in Qualitative Rehabilitation Research. *Qual Health Res* 2017;27.
- Moberg J, Alonso-Coello P, Oxman A. GRADE Working Group. GRADE Evidence to Decision (EtD) Frameworks Guidance. Version 1.1. 2015 <https://ietd.epistemonikos.org/#/help/guidance>.
- Verkerk K, Van Veenendaal H, Severens JL, et al. Considered judgement in evidence-based guideline development. *Int J Qual Health Care* 2006;18:365–9.
- Moreira T. Diversity in clinical guidelines: the role of repertoires of evaluation. *Soc Sci Med* 2005;60:1975–85.
- Calderón C, Rotaecche R, Etxebarria A, et al. Gaining insight into the Clinical Practice Guideline development processes: qualitative study in a workshop to implement the GRADE proposal in Spain. *BMC Health Serv Res* 2006;6:138.
- Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. *BMC Med Res Methodol* 2009;9:59.
- Lomas J. Words without action? The production, dissemination, and impact of consensus recommendations. *Annu Rev Public Health* 1991;12:41–65.
- Peirce C. *Philosophical Writings*: Dover Publications Inc, 1986.
- Douven I. Abduction. *The Stanford Encyclopedia of Philosophy* 2011. <https://plato.stanford.edu/archives/sum2017/entries/abduction/>
- Popper KR. *Objective Knowledge: An Evolutionary Approach*: Clarendon Press, 1979.
- Kriebel D, Tickner J, Epstein P, et al. The precautionary principle in environmental science. *Environ Health Perspect* 2001;109:871–6.
- Pollock JL. *The Logical Foundations of Means End Reasoning. Common Sense, Reasoning, and Rationality*. Oxford: Oxford University Press, 2002:60–77.
- Stolper E, Van de Wiel M, Van Royen P, et al. Gut feelings as a third track in general practitioners' diagnostic reasoning. *J Gen Intern Med* 2011;26:197–203.
- Gigerenzer G, Brighton H. Homo heuristicus: why biased minds make better inferences. *Top Cogn Sci* 2009;1:107–43.