



# Wrong guidelines: how to detect them and what to do in the case of flawed recommendations

Primiano Iannone,<sup>1</sup> Giorgio Costantino,<sup>2</sup> Nicola Montano,<sup>2</sup> Gian Marco Podda,<sup>3</sup> Monica Minardi,<sup>4</sup> James Doyle,<sup>4</sup> Antonino Cartabellotta<sup>5</sup>

10.1136/ebmed-2016-110607

<sup>1</sup>Emergency Department, Ospedale del Tigullio, Lavagna, Italy

<sup>2</sup>Fondazione Ca' Granda IRCCS Ospedale Maggiore Policlinico Dipartimento di Scienze Cliniche e di Comunità Università degli Studi di Milano, Milano, Italy

<sup>3</sup>U.O. Medicina III Ospedale San Paolo, Milano, Italy

<sup>4</sup>Emergency and Intensive Care Department, Royal Surrey County Hospital NHS Foundation Trust, Guildford, UK

<sup>5</sup>GIMBE Foundation, Bologna, Italy

Correspondence to  
**Dr Primiano Iannone**,  
Emergency Department,  
Ospedale del Tigullio,  
via Don Bobbio 25,  
Lavagna 16033, Italy

## Abstract

Any evidence-based recommendation needs careful assessment of its methodological background as well as of its content trustworthiness, especially given that following it will not necessarily produce the intended clinical outcomes. There are no established instruments to evaluate guidelines for their content, while useful tools assessing the quality of methods followed are well recognised and adopted. We suggest a 'safety bundle' considering methodological aspects and content trustworthiness of guidelines, by adopting the GRADE method in a backward fashion. Sharing the critical analysis of the guidelines with patients, including any eventual uncertainty about them, is of key importance in order to avoid the possible adverse effects derived from following the wrong guidelines. Such critical approach is also helpful and beneficial in producing better care pathways, health policy decisions and more relevant and ethical research.

## Introduction

In the first article, we described the potential flaws often undermining guideline trustworthiness, despite their evidence-based 'quality mark', and mentioned the considerable efforts made to counter this over the past two decades. We have highlighted that limited panel composition, poor methodology and conflict of interest are the main 'endogenous' reasons for their weakness to date. Moreover, the 'corruption' of primary evidence and waste of biomedical research on which they are grounded further contributes to that phenomenon.<sup>1</sup>

Once published, it is hard to discredit even the most conflicted and untrustworthy guideline,<sup>2</sup> mainly for the asymmetry of biomedical literature, not giving adequate resonance to such errors, inertia of medical practice, fear of medicolegal claims and the absence of independent, trustable and authoritative agencies rating, not simply collecting, the guidelines. Therefore, it is crucial not only to inform clinicians that guidelines can be wrong much more commonly than expected, but also to explain to them how to recognise the flaws and errors confidently.

In this paper, we suggest how to achieve this. A broader perspective by the research and policymaking standpoint about the issue will also be offered.

## Assessing the trustworthiness of a recommendation

Most physicians need to decide about the trustworthiness of a single recommendation, not of an entire guideline. In doing so, while the above-cited extrinsic factors

limiting the reliability of evidence-based guidelines need to be always considered, the conceptualisation of trustworthiness as related to its content and/or methodological value is very helpful, together with the need of deciding a 'trustworthiness threshold' to be adopted, depending on the clinical circumstances encountered (figure 1).

Indeed, assessing the quality of the methods followed to produce the guideline containing that recommendation is a reasonable, preliminary and necessary step. Up to 22 analytical tools exploring the quality of guidelines have been set up through 2005<sup>3</sup> and other frameworks appeared later or were significantly amended. Some of these tools gained large popularity and most literature assessing quality of guidelines has adopted one of these instruments.<sup>4-6</sup> They share similar quality domains and the items to be assessed overlap to a large extent (table 1).

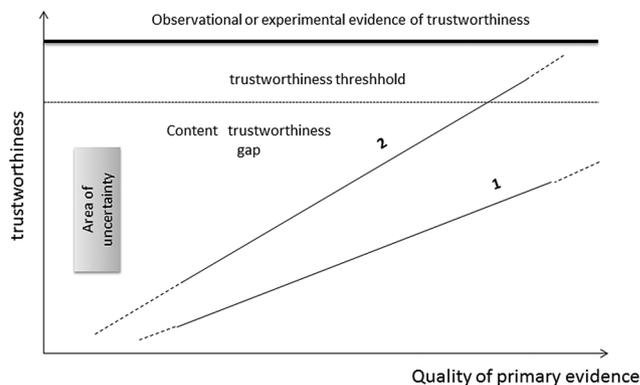
However, 'sensitivity' and 'specificity' of these tools have never been assessed prospectively, they have been decided mostly on a consensus basis, and they do not offer a score with a clear cut-off value below which a guideline should be disregarded. The most recent of them, from Institute of Medicine, sets the standard to such a high level that it is unfulfilled by all guidelines so far measured against it,<sup>7</sup> so its practical value is unclear.

Alternatively, 'red flags' with an empirical relationship to flawed recommendations have been proposed,<sup>8</sup> but they lack external validation and it is unlikely that such a knowledge gap will be reduced soon. Guideline repositories such as the National Guideline Clearinghouse have recently set forth a minimum standard of inclusion criteria for guidelines to be admitted and accepted onto that database. Alternatively, Guidelines International Network advocates and encourages public involvement in evaluation and discussion of guidelines, whereas NICE offers a 'methodological seal' for those organisations producing guidelines on its behalf through the NICE accreditation programme. Effectiveness of these systems in orienting readers towards the most trustworthy guidelines is however unknown.

Nevertheless, the lower the methodological quality of a given guideline, the higher the need to assess its content trustworthiness before adopting that recommendation with confidence, especially in the case of high risks of harm to the patients and/or high costs deriving from its adoption. Therefore, a complementary and even closer view must be obtained considering the 'content' trustworthiness, that is how and to what extent, the evidence was correctly rated and translated onto the recommendation of interest. When more than one guideline



► <http://dx.doi.org/10.1136/ebmed-2016-110606>



**Figure 1** 1: Low-quality guidelines; 2: high-quality guidelines. The tradeoff between quality of guideline methods, evidence available and trustworthiness. Decision threshold of trustworthiness has to be decided on a case-by-case basis. When little or no evidence is available (area of uncertainty) and recommendations follow from a consensus process, trustworthiness can be achieved only by assessing outcomes prospectively.

on the same topic exists, searching for any unexplained difference between recommendations considering the same evidence is a popular, indirect ('proxy') index of trustworthiness and reasoned comparisons of similar guidelines are now offered by the National Guideline Clearinghouse website. Unfortunately, the concordance of recommendations between guidelines may be falsely reassuring and ultimately wrong,<sup>9</sup> while discordance may be (legitimately) due to differing interpretations of the 'evidence', the relative importance attributed to outcomes considered or to variable context-sensitive issues.<sup>10</sup> Eventually, discrepant and discordant recommendations lead to further confusion. Instead, assessing indepth the content trustworthiness of recommendation implies taking into account, among the others, the consistency of the recommendation with the primary evidence on which it was grounded, how evidence gaps were filled-in by the discussion within the panel group, what values and considerations were adopted, and from which perspective the recommendation was issued (individual clinical practice or population wide/policy-making) and whether costs were considered or not.

The complexity of factors involved in producing recommendations informed by—rather than simply based on—evidence has been made explicit by the GRADE working group,<sup>11</sup> which produced a method increasingly adopted (although not universally) by many organisations and societies sponsoring guidelines. Also, the GRADE method can be conveniently used by guideline users needing to scrutinise the trustworthiness of a given recommendation indepth (see table 1).

'Cross matching' a recommendation with the related primary evidence using the GRADE method with a 'backward' reanalysis going from the recommendation to the evidence which produced it is the essence of this 'direct' approach. However, this method is extremely time-consuming, requires considerable expertise of evidence-based methodology and its reproducibility is unknown. Furthermore, this approach has been evaluated only in a few studies.<sup>10 12</sup> On the contrary, content (and methodological) trustworthiness can be verified also through a more affordable assessment of literature

discussing guidelines and primary studies to which they refer to, as well as following the debate on websites dedicated to guidelines, letters, editorial and comments in peer-reviewed journals, blogs, social media and so on, if sufficient time (and interest) is dedicated to it. The highest level of content trustworthiness can be achieved only when prospective, convincing observational and/or experimental data about the outcomes derived from the adoption of a given recommendation is available ('ex post trustworthiness'). However, this knowledge is rarely available.<sup>2 13–15</sup> Furthermore, there is no structured effort made by anyone (either at an organisational or individual level) to assess systematically the outcomes derived from adopting guidelines. Eventually, 'modelling' of the projected health outcomes seems necessary,<sup>16</sup> and this is increasingly considered a good option for screening programmes and population-wide interventions.

In conclusion, assessing the reliability of guidelines is largely a matter of individual, careful and complex judgement. Tools assessing their methodology are helpful but not exhaustive, and no predefined approach can be adopted to assess their content trustworthiness, especially when robust data are lacking.

### Building a guideline user safety bundle

Since there is no single easily distinguishing feature of a trustworthy guideline and no satisfactory comprehensive assessment tool available, we suggest for the guideline user a seven-point 'safety bundle', to recognise wrong guidelines with confidence.

- 1 *Question context:* First of all, framing the question correctly is of paramount importance. To this regard, it must be clear what reason that recommendation is searched for. That is whether we need to be answering individual patient problems or we are in search for answers to effectiveness, safety or cost issues for population-wide interventions, setting a standard for quality assessment purposes or deciding how to allocate resources properly. Regardless, a clear translation of the generic question into an answerable

**Table 1** Analytical tools useful for assessing guidelines' trustworthiness

Methodological trustworthiness	Quality domains	Items	Website/reference
AGREE	1 Scope and purpose 2 Stakeholder involvement 3 Rigour of development 4 Clarity of presentation 5 Applicability 6 Editorial independence	23	<a href="http://www.agreetrust.org">http://www.agreetrust.org</a>
GIN	1 Composition of guideline development group 2 Decision-making process 3 Conflicts of interest 4 Scope of a guideline 5 Methods 6 Evidence reviews 7 Guideline recommendations 8 Rating of evidence and recommendations 9 Peer review and stakeholder consultations 10 Guideline expiration and updating 11 Financial support and sponsoring organisation	–	<a href="http://www.g-i-n.net/">http://www.g-i-n.net/</a>
IOM	1 Establishing transparency 2 Management of conflict of interest (COI) 3 Guideline development group composition 4 Clinical practice Guideline–systematic review intersection 5 Establishing evidence foundations for and rating strength of recommendations 6 Articulation of recommendations 7 External review 8 Updating	20	Ref 6
Lenzer 'red flags'	1 Conflict of interest 2 Panel composition 3 External review	3	Ref 8
Content trustworthiness			
GRADE	1 Quality of evidence a Relevance of outcomes b Consistency c Precision d Risk of bias e Publication bias f Confounders g Dose-gradient effect 2 Balance of favourable and unfavourable outcomes 3 Patients values and preferences 4 Costs/use of resources	–	<a href="http://www.gradeworkinggroup.org/">http://www.gradeworkinggroup.org/</a>

- clinical question is required, according to the most basic and time-honoured evidence-based rule (remembering the PICO acronym for Patients, Intervention, Comparison and Outcomes).
- Trustworthiness threshold:** Then, we should decide the 'trustworthiness' threshold for the recommendation we will eventually find, depending on the relevance of the outcomes we are considering, the expected benefits, the costs derived from adopting that recommendation, the potential risks and all of the patients' (or other stakeholders) perspectives appropriate for that question.
- Remit and scope:** Considering the remit and scope of the guideline we are considering is worthy of attention too. Some guidelines, for example, may include cost considerations and offer a population-wide perspective, whereas others offer a more patient-centred approach. It would therefore be of use to find a guideline tailored to the perspective which is more appropriate to our needs.
- Quality assessment:** Using a quality assessment tool such as AGREE, IOM or GIN criteria will increase the users' awareness of the potential pitfalls of a guideline. There is empirical evidence that non-adherence to some key features of the guideline making process increases the risks of wrong recommendations. Among them, conflict of interest, panel composition, poor and opaque methods of evidence rating and absence of external review process will all erode substantially the reliability of guidelines and can be used as screening tests of methodological trustworthiness. Since medical specialty guidelines often fall short of these basic standards, they must be considered with special attention.
- Content assessment:** An assessment of recommendations against their content trustworthiness is also necessary, especially when there is any reasonable doubt about their methodological quality, when the risks and/or costs deriving from their adoption are of special concern, when there is weak or conflicting

evidence available or a high 'trustworthiness threshold' is considered appropriate altogether. We discourage the practice of cross matching similar guidelines as a surrogate for their content trustworthiness. Instead, we recommend to control accurately whether external lines of evidence support that recommendation, outcomes following the adoption of that recommendation are available, as well as to use the GRADE method (or at least using broadly the GRADE principles of evidence rating) to scrutinise their content trustworthiness. It should be remembered again that the 'corruption of the evidence' makes any effort to produce reliable and useable recommendations difficult to overcome.

- 6 **Adoption:** Once a reasonable level of certainty about the reliability of that recommendation has been achieved, the next step is to decide whether it is worth adopting it as is, modifying it or disregarding it on a case-by-case basis. The evidence to decision frameworks of the DECIDE project (<http://www.decide-collaboration.eu/>) may be helpful for this task.
- 7 **Shared decision:** The careful critical assessment of any recommendation leading to this analysis should not be used for solitary decisions but has to be communicated effectively and explicitly with the patient for a shared decision-making process.

## Conclusions

Evidence-based guidelines are a valuable tool in many cases for most physicians willing to adopt the best biomedical knowledge available to improve patients' care. In the last years, attention has been paid to translate accurately the general purpose of guidelines to the individual patients' needs and specific frameworks have been developed for making this process explicit, reasonable and consistent with the available evidence.<sup>17</sup>

Despite their evidence-based quality mark, however, even contemporary guidelines may be flawed, not only for internal biases but also for the wider 'corruption' and waste of the biomedical research under the pressure of commercial interests, excess medicalisation, overtreatment and defensive medicine.

Dealing with flawed guidelines is not rare and we need to assess their trustworthiness more firmly. Unfortunately, the process which leads to the right decision is not straightforward, because of the high number of flawed guidelines. We suggest the adoption of a guideline user 'safety bundle' which considers all of the intrinsic and extrinsic reasons of guideline untrustworthiness, to make an informed and cautious use of guidelines before adopting (and/or adapting) them with confidence. We also strongly encourage the use of effective and explicit communication with the patient, to foster a comprehensive shared decision-making process.<sup>18 19</sup>

Beside the individual patient level, untrustworthy guidelines raise serious concerns at the level of team-work practice, design and implementation of care pathways as well as in defining correctly the research priority agenda, not to mention their medicolegal implications. Eventually, discovering a 'suspicious' recommendation may be a valuable opportunity for professional and

scientific growth, and a discussion about a valuable opportunity for reflecting on care pathway production.

Moreover, in an ideal and healthy scientific community, finding an untrustworthy recommendation should elicit an indepth, unconflicted reassessment of that guideline on an urgent basis, considering the huge potential negative impact on patient's safety, the harms and wastes from a wider societal perspective and the devastating consequences derived from the assumption of false equipoise based on flawed guidelines. Furthermore, there is the potential to create a serious obstacle to otherwise necessary research.<sup>20</sup>

The scientific organisation producing that suspicious guideline should be notified and medicolegal value of that guideline downgraded. Peer-reviewed journals should be more involved in this critical appraisal process. If concerns are justified, adequate explanations of the 'critical incident' should be offered by guideline sponsors and expert panelists, methods of guideline making process eventually reassessed and that recommendation amended expeditiously.

As the editor of *Lancet* Richard Horton said, guidelines (should) 'force us to scrutinise primary research literature in ways that we don't normally do'. We agree. That is exactly the opposite of using them as heuristic tricks to respond quickly to complex clinical questions needing a slow, rather than fast, decision-making process.<sup>21</sup>

Despite the optimistic expectations announced more than two decades ago, primary research and its 'evidence based' synthesis offered by guidelines can be unreliable, flawed and untrustworthy. However, thanks to evidence-based methods, such problems have been highlighted and a 'methodological immunisation'<sup>22</sup> from wrong guidelines and 'bad research'<sup>23</sup> offered as a solution.

Meanwhile, a critical appraisal of the primary evidence the guidelines refer to and considering to what extent they are trustworthy and relevant for the specific clinical context remain the sole and ultimate way of protecting our patients from the risks and harms caused by wrong guidelines.

Twitter Follow Primiano Iannone @primianoianone

**Contributors** PI contributed to conception and design of the paper, stewardship of the discussion within the panel of the authors for the production of the safety bundle, critical assessment of the literature about guideline methodology and appraisal, selection of case studies cited; draft of the manuscript. NM, MM and AC made substantial contributions to the acquisition, analysis and interpretation of current methodology of guidelines appraisal; revision of the manuscript for important intellectual content; final approval of the version published. JD, GC and GMP made substantial contributions to the analysis, interpretation of relevant literature considered and of the case studies to be tested with the safety bundle; revision of the manuscript for important intellectual content; final approval of the version published. All of the authors agree 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.

**Competing interests** None declared.

**Provenance and peer review** Not commissioned; internally peer reviewed.

**References**

1. Iannone P, Montano N, Minardi M, *et al.* Wrong guidelines. Why and how often they occur. *Evid Based Med* 2017;22:1–3.
2. Graham DC, Darrel PF. Perioperative  $\beta$  blockade: guidelines do not reflect the problems with the evidence from the DECREASE trials. *BMJ* 2014;349:g5210.
3. Vlayen J, Aertgeerts B, Hannes K, *et al.* A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *Int J Qual Health Care* 2005;17:235–42.
4. Qaseem A, Forland F, Macbeth F, *et al.* Guidelines international network: toward international standards for clinical practice guidelines. *Ann Intern Med* 2012;156:525–31.
5. Brouwers MC, Kho ME, Browman GP, *et al.* AGREE Next Steps Consortium. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ* 2010;182:E839–42.
6. Institute of Medicine. *Clinical practice guidelines we can trust*. Washington, DC: The National Academies Press, 2011.
7. Ransohoff DF, Pignone M, Sox HC. How to decide whether a clinical practice guideline is trustworthy. *JAMA* 2013;309:139–40.
8. Lenzer J, Hoffman JR, Furberg CD, *et al.*, Guideline Panel Review Working Group. Ensuring the integrity of clinical practice guidelines: a tool for protecting patients. *BMJ* 2013;347:f5535.
9. Iannone P, Haupt E, Flego G, *et al.* Dronedronone for atrial fibrillation: the limited reliability of clinical practice guidelines. *JAMA Intern Med* 2014;174:625–9.
10. Eisinger F, Geller G, Burke W, *et al.* Cultural basis for differences between U.S. and French clinical recommendations for women at increased risk of breast and ovarian cancer. *Lancet* 1999;353:919–20.
11. GRADE working group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490–4.
12. Paul E, Alexander PE, Brito JP, *et al.* World Health Organization strong recommendations based on low-quality evidence (study quality) are frequent and often inconsistent with GRADE guidance. *J Clin Epidemiol* 2016;72:98–106.
13. Gerstein HC, Miller ME, Byinton RP, *et al.*, Action to Control Cardiovascular Risk in Diabetes Study Group. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med* 2008;358:2545–59.
14. Hsieh YH, Rothman RE, Laeyendecker OB, *et al.* Evaluation of the Centers for Disease Control and Prevention recommendations for hepatitis C virus testing in an urban emergency department. *Clin Infect Dis* 2016;62:1059–65.
15. Gilbert R, Salanti G, Harden M, *et al.* Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of recommendations from 1940 to 2002. *Int J Epidemiol* 2005;34:874–87.
16. U.S. Preventive Services Task force. Use of decision models in the development of evidence-based clinical preventive services recommendations: methods of the U.S. Preventive Services Task Force. *Ann Intern Med* 2016;165:501–8.
17. Alonso-Coello P, Schünemann HJ, Moberg J, GRADE working group. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ* 2016;353:i2016.
18. Hoffmann TC, Montori VM, Del Mar C. The connection between evidence-based medicine and shared decision making. *JAMA* 2014;312:1295–6.
19. Yudkin JS, Kavanagh J, James P, *et al.* Guidelines for treating risk factors should include tools for shared decision making. *BMJ* 2016;353:i3147.
20. Kavanagh BP. The GRADE system for rating clinical guidelines. *PLoS Med* 2009;6:e1000094.
21. Kahneman D. *Thinking, fast and slow*. The New York Time Publisher. 2011.
22. Liberati A. *Etica, Conoscenza e Sanità*. Il Pensiero Scientifico Editore. 2005:3–37.
23. Ioannidis JPA. Why most clinical research is not useful. *PLoS Med* 2016;13:e1002049.