Mitigating intersubjective variability in clinical guidelines

Henock G Yebyo, PhD[1], Milo A Puhan, MD, PhD[1], Gerben ter Riet, MD, PhD[2,3]

[1] Department of Epidemiology, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, CH-8001 Zurich, Switzerland
[2] Urban Vitality Centre of Expertise, Faculty of Health, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands
[3] Department of Cardiology, Amsterdam University Medical Center (Amsterdam UMC) Location AMC, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

World count: 1089

We share the worries Bauchner and Ioannidis expressed in their recent viewpoint on *The Subjective Interpretation of the Medical Evidence,*[1] in which they identify conflicts of interest (COI), evidence selection, and evidence appraisal as the among the main culprits of difficult-to-understand variability in clinical guideline recommendations. They end by saying that the public should understand that, even though uncertainty is common, it should not paralyze us, and medical decisions can be made based on best evidence "once values, preferences, and biases are vetted.".

We think it is prudent to accept that in the assessment of scientific evidence elements of subjectivity will remain, and that the creation of clinical practice guidelines is a social endeavour. Therefore, it is useful to strive for intersubjectivity supported by well-designed processes that achieve ideal speech situations, thus ensuring inclusivity, equality, truthfulness, and rationality.[2]

While we briefly discuss meticulous development of clinical guidelines, which is much discussed,[1] here we emphasize the importance of integrating formal benefit-harm analysis (BHA) to mitigate the limitations in intersubjective interpretations of evidence. We suggest how incorporating patient preferences, employing sensitivity analysis techniques such as 'vibration of effects' [3] and 'multiverse analysis', and ensuring transparency in methods and data have the potential to importantly reduce variability in clinical guideline recommendations and shed light on all the sources of any remaining variability.

Once a theme for a clinical guideline has been identified, a guideline panel must be formed. The Guidelines International Network (GIN) provides excellent guidance on how to do this (https://g-i-n.net). A balanced and inclusive selection of committee members is crucial to balancing any biases that cannot be fully excluded. Traversy et al. propose the following approach for panel formation that we support: "*to find panelists without COIs, developers might consider broadening their search beyond their usual pool of candidates to include those at earlier career stages, from other clinical areas, or even from other disciplines*."[2] All candidates are required to be fully open about their COIs or issues that are likely to be perceived as such and be held accountable for full reporting these. Clear criteria should be in place that define the threshold above which membership is impossible. Already, such procedures are often in place, but the devil is in the details: who determines

candidates for panel membership? What happens if candidates omit particular ties with important stakeholders? Traversy et al. cite French law that allows for up to 5 years of imprisonment and up to a US$558,000 fine for non-disclosure (www.has-sante.fr).

After identifying the clinical issues for recommendations, it is crucial to establish criteria for the collection, appraisal and synthesis of evidence. Methods like those of the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) group are commonly used to streamline these processes (https://www.gradeworkinggroup.org). However, these approaches resolve differences that inevitably arise during the review of evidence by various stakeholders through consensus. As Bauchner and Ioannidis also suggest, this approach lacks reproducibility and measurability. To address this, more systematic and quantitative methods are required.

BHA is a method that provides a structured and quantitative framework for simultaneously evaluating both positive and negative outcomes.[4] By incorporating key decision-making parameters such as treatment effects, risk distributions, and patient preferences, a base-case BHA enables the development of tailored recommendations that are likely to achieve a net benefit for individuals, subgroups, or populations. It is not uncommon for some guidelines to set thresholds for interventions but often without full transparency or systematic evaluation. For instance, statins are usually prescribed when the 10-year cumulative cardiovascular disease (CVD) risk is 7.5% (or above depending on the guideline),[5] but the method behind this threshold often lacks clarity and fails to reflect a true benefit-harm evaluation. BHA enables the determination, on a common scale, of the thresholds for initiating statins by weighing the benefits in preventing CVD against potential risks, including muscle issues, diabetes, and liver or kidney dysfunctions.[6]

Moreover, BHA's versatility extends to the exploration of alternative assumptions in decision-making processes, beyond objective evidence. It accounts for both statistical and non-statistical uncertainties, varying outcome risks and 'vibration of effects',[3] alternative strategies, as well as subjective factors like the opinions of stakeholders, physicians, and patients. This consideration is particularly important in the development of clinical guidelines where most source studies are reductionist and may not fully capture the complexity of real-world scenarios. In such situations, incorporating expert opinions, which provide additional perspectives by reflecting diverse experiences and observations, is essential. BHA entertains multiverse analysis, a form of radical sensitivity analysis that examines the above multiple alternative perspectives and courses of actions for their impact. This approach fosters the development of more comprehensive and inclusive guidelines, ensuring the best possible patient outcomes.

It is often said that science is about the state of affairs, the *is*, but cannot say anything about the *ought* or *should*. We agree, but also think that once we agree on valid ways to derive patient preferences, BHA allows us to incorporate these preferences that reflect patients' values, their willingness to accept harms in pursuit of the treatment benefit, and influences of treatment outcomes on their daily lives. While, admittedly, some BHA-based guideline recommendations may involve using average preferences in clinically defined patient subgroups, many decisions are preference-sensitive. Large variability in patient preferences further challenges the current application of aggregated recommendations, underscoring the need for a tailored approach to individuals. Similarly, the risks of both positive and negative treatment outcomes vary across individual patients or subgroups. BHA offers a framework that supports personalized decision-making by integrating these individualized factors.

Transparency of methods and data is a key component of approaching ideal speech situations in guideline development.[7] Transparency involves the following components: firstly, procedures for the selection into guideline panels. Secondly, the selection of evidence that will (not) be looked at by the committee and by which methods. Here radical sensitivity analysis is also recommended, where the BHA approach would be instrumental to systematically examine alternative paths. Thirdly, transparent analysis and explanation, ideally by an independent third party, why certain recommendations differ from those in an existing guideline on the same topic, and if needed, stating that it is unclear why certain recommendations differ. And finally, sharing of protocols, code, logs with key decisions during meetings, and data.

In conclusion, we agree with Bauchner and Ioannidis that medical decisions can be made based on the best evidence once values, preferences, and biases are vetted. We emphasize the importance of systematic BHA to integrate evidence, preferences, and intersubjective variations and to formally evaluate their impact. This approach enhances the transparency and personalization of clinical guidelines.

References

1. Bauchner H, Ioannidis JPA. The Subjective Interpretation of the Medical Evidence. *JAMA Health Forum*. 2024;5(3):E240213. doi:10.1001/jamahealthforum.2024.0213
2. Traversy G, Barnieh L, Akl EA, et al. Managing conflicts of interest in the development of health guidelines. *CMAJ*. 2021;193(2). doi:10.1503/cmaj.200651
3. Ioannidis JPA. Why most discovered true associations are inflated. *Epidemiology*. 2008;19(5). doi:10.1097/EDE.0b013e31818131e7
4. Puhan MA, Singh S, Weiss CO, Varadhan R, Boyd CM. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. *BMC Med Res Methodol*. 2012;12:1-12.
5. Mortensen MB, Nordestgaard BG. Comparison of five major guidelines for statin use in primary prevention in a contemporary general population. *Ann Intern Med*. 2018;168(2):85.
6. Yebyo HG, Aschmann HE, Puhan MA. Finding the balance between benefits and harms when using statins for primary prevention of cardiovascular disease: A modeling study. *Ann Intern Med*. 2019;170(1). doi:10.7326/M18-1279
7. Hofmann W. Jürgen Habermas: Theorie des kommunikativen Handelns. Bd. 1: Handlungsrationalität und gesellschaftliche Rationalisierung; Bd. 2: Zur Kritik der funktionalistischen Vernunft, Suhrkamp: Frankfurt 1981, 533 + 632 S. In: *Klassiker Der Sozialwissenschaften*. 2016. doi:10.1007/978-3-658-13213-2_74