Experts often subjectively disagree on how they interpret the same evidence and what recommendations they derive from it.¹ Meticulous processes to resolve diverging views in guideline development efforts, for example, may not remove subjectivity. Even the most prestigious organizations sometimes have different guideline recommendations. Subjective disagreements can be common, extreme, and unsettling when evidence is limited and rapidly evolving—as in many questions related to COVID-19. However, subjectivity exists, and differences ensue even for common diseases where evidence has accrued and been evaluated for decades. For example, the American College of Physicians, the American Cancer Society, and the US Preventive Services Task Force (USPSTF) vary on when to initiate screening for colorectal cancer and the preferred screening methods.²⁻⁴ Breast cancer and depression screening recommendations have been debated for decades.

Recommendations are naturally expected to evolve as new evidence develops. However, different experts and guideline committees often have different selection criteria and judgments about what new evidence is worth considering and how valuable it may be. For example, whereas the USPSTF does not recommend use of the coronary artery calcium score to evaluate cardiovascular risk, numerous guidelines from professional cardiology societies, such as the American College of Cardiology/American Heart Association, do recommend it. Most USPSTF members are generalists, whereas most members of guideline committees of specialty societies are specialists. Specialists may view the same evidence differently than generalists do.

Experts may also frame recommendations differently if they know that their statements will affect insurance coverage or actual policy. The Affordable Care Act mandates that A- or B-level screening recommendations from the USPSTF must be covered by private insurers without costs for patients. USPSTF members are certainly aware of the importance of their recommendations with respect to insurance coverage. They could be influenced by this legal mandate to avoid an “I statement” (ie, insufficient evidence, insurance coverage not required).

Different guideline development methods may lead to different recommendations. For example, the USPSTF follows a well-documented approach, including an independent data synthesis produced by an evidence center. Many guidelines follow the popular Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach. Validation data for each separate step of the process suggest relatively good reproducibility. However, the sum of variability across all steps can become substantial⁵ and is poorly understood. Groups of experts may frame the relevant questions differently, select different data to be included, and weigh aspects of the evidence differently. For example, the USPSTF relies on data from randomized clinical trials (RCTs) and de-emphasizes modeling studies. Conversely, a modeling study, in part, led to a change in the recent lung cancer screening guideline of the American Cancer Society.⁶

Specialist societies and advocacy groups, particularly those representing patients, receive substantial support from industry, which may influence their recommendations. However, even without financial conflicts, allegiance bias and ideology may also shape recommendations, especially for contentious topics such as COVID-19, cancer screening, or nutrition. Regardless of financial conflicts of interest, patient and professional groups may include members who are known to have strong beliefs and sometimes express overt advocacy and passionate activism.
Across most medical interventions, evidence is often so weak that recommendations largely depend on subjective opinions. Recommendations may also be affected not only by the evidence on specific interventions, but also by what other interventions are available. For example, with respect to the pharmaceutical treatment of obesity in adolescents, the American Academy of Pediatrics recently gave it a B recommendation, despite limited evidence. Perhaps members of the committee were influenced by the increasing prevalence of obesity and the lack of effective behavioral treatment for obesity in adolescents.

Besides organized guideline committees, individual experts have been increasingly influential and vocal in the public space. Many medical experts have become powerful influencers in media and on social media, interpreting evidence subjectively. Journalists and other nonprofessional influencers who may lack an understanding of how evidence is developed and evaluated nevertheless join the ensuing discussions and debates. During dramatic debates about COVID-19 policies, gender-affirming care, and the pharmaceutical treatment of obesity in children, the subjective interpretation of evidence has emerged as an important issue for medical reporters and medical experts.

There has been substantial concern about misinformation and disinformation in the reporting of medical news. There is increasing public perception that reporting is not objective, but rather selective and subjective, even from otherwise most-respectable venues. Hyperbolic claims for major discoveries and exaggerated benefits of both mainstream and controversial treatments (“cures”) are common. Moreover, adoption of polarized, advocacy-driven worldviews are probably contributing to public loss of trust in medical science and public health.

The reporting of results from RCTs is a good example. Even in a positive RCT, the number needed to treat can be large; that is, most individuals will not benefit. Additional nuances may exist but may be missed in the news reporting (eg, the difference between clinically important vs statistically significant effects, the intricacies in interpreting noninferiority results, the challenges of understanding the results of composite end points with multiple components). Investigators, journal editors, and those who report the news or are interviewed as experts should clearly express this uncertainty and its nuances, as well as the amount and quality of evidence that supports their statements.

Subjective interpretation of evidence is complex and a challenging problem to mitigate or solve. Interventions to diminish it should be formally tested. Researchers should adhere to the primary objective and findings of studies, avoiding hype. Registration has improved the reporting of RCTs and meta-analyses and could be extended to more study types. Editors can be gatekeepers to ensure that authors use appropriate language (eg, causal vs association, clinical vs statistical significance) and include the number needed to treat for RCTs. Recommendations from professional groups will probably continue to diverge. What is critical is that they make clear by whom and how the recommendations were developed and minimize financial or other conflicts of interest. Guideline committees could also consider having a section in which they explain how and why their guideline differs from other guidelines on the same topic—a practice that the USPSTF has already adopted.

Finally, individuals who write about medicine or are quoted as experts should be aware of their own biases, reflecting on whether they have assessed the evidence impartially. Given that self-reflection is not easy, editors and media hosts should assess bias in their commentators. Unfortunately, even editors and hosts themselves may be biased, even more than their commentators. Therefore, the public should be reminded that uncertainty—and even the extent of uncertainty—is common. This realization is not paralyzing; medical decisions can still be made with the best available evidence, once values, preferences, and biases are carefully vetted.
REFERENCES


