modified AGREE and using systematic reviews to supplement the evidence base when there is discordance among recommendations.

**Methods** This AHRQ funded work used a mixed-methods approach. We performed semi-structured telephone interviews and surveys to query developers about impressions of and intentions to implement the IOM standards in their CPGs. We also performed our own assessments of guidelines and compared them with developer self-ratings.

**Results** Of 14 developers, 43% utilised a systematic review to underpin their guidelines, and 57% felt they would in the future. Funding sources were not disclosed by 46% of the developers. While 80% utilised an evidence rating scheme, fewer rated the recommendations. Notable differences between developer self-ratings and researcher assessments of adherence occurred in several areas.

**Discussion** While some developers intend to improve processes to meet the Standards, others acknowledged they will not. Yet still others felt they already met the standards, but our assessment suggested a different estimation, revealing varying understanding among developers of the Standards.

**Implications for Guideline Developers/Users** The IOM standards will help identify rigorous and transparent evidence-based guidelines, but will pose implementation challenges. Education of developers on the Standards and expectations around them will be critical.

**Guideline Developers’ Self-Perceptions of Adherence to and Intentions to Adhere to the IOM Standards**

**Objectives** (1) Assess developers’ self-perceptions of adherence to the IOM standards (2) Assess developers’ intentions to adhere to the IOM standards.

**Background** The 2011 IOM report called for more rigorous and transparent development of guidelines. Compliance with the IOM Standards may be challenging for developers. Developer perception of their current adherence to the Standards gives insight into their understanding of them and the likelihood of adhering to them in the future.

**Methods** We examined 17 international, English-language guideline development systems to identify rating parameters and applied descriptive statistics. We also searched for conceptual linkages in the rating system descriptions and identified systems where Strength of Recommendation was stated as an Evidence Quality parameter.

**Results** Rating systems were remarkably inconsistent in their application of category indicators—using letters, Arabic and Roman numerals and combinations. The modal and median number of Evidence Quality categories was 3 (range 0 to 10) and Recommendation Strength categories was 4 (range 0 to 6). 13/17 used randomised trials as indicators of highest quality evidence. 7 systems used “expert opinion,” 6 used “case report” or “case series,” and 4 described “reasoning from first principles” to define lowest evidence quality. Definitions of intermediate levels varied considerably. 7 systems judged benefits and harms in deriving Recommendation Strength. In 7 rating systems, Strength of Recommendation was described entirely in terms of Evidence Quality.

**Implication** There is considerable disagreement about the requisite granularity and definition of categories of Evidence Quality and Recommendation Strength. Application of the concept of Recommendation Strength consonant with the IOM standard is limited. A straightforward mapping of rating systems to one another is elusive.