

Assessing Quality in Systematic Reviews: Application of AMSTAR Criteria to Plastic Surgery Literature

Hillary E. Jenny, MPH; Benjamin B. Massenburg, BA; Joseph Leanza, BA, Peter J. Taub, MD

Introduction:

As medicine becomes increasingly evidence-based, it is crucial to understand the level of evidence informing healthcare decisions. Systematic reviews are thought to provide high-quality evidence, but they vary markedly in methodological rigor. A Measurement Tool to Assess systematic Reviews (AMSTAR) was created in 2007 to evaluate scientific quality using an 11-item scale. Studies are classified as high (scoring 8-11), medium (4-7), or low quality (0-3). The following investigation uses AMSTAR criteria to assess the quality of systematic reviews in the plastic surgery literature.

Methods:

PubMed and Web of Science electronic databases were searched for systematic reviews of diagnostic or management interventions published from January 2000-August 2015 in the 13 highest-impact plastic surgery journals. Narrative reviews, discussion articles, reviews without transparent search criteria, and reviews of disease prevalence or research methodology were excluded. Two reviewers (HEJ, BBM, or JL) independently screened titles and abstracts for inclusion. Full texts for the included abstracts were screened, and included studies were rated according to AMSTAR guidelines. Discrepancies were resolved through discussion. Mean AMSTAR score was calculated for each included journal with ≥ 1 included review, ANOVA compared each journal's mean AMSTAR score, and an independent-sample T-test compared mean AMSTAR scores prior to and after AMSTAR creation.

Results:

The database search identified 807 non-duplicate reviews. 410 were excluded on title and abstract review and 92 on manuscript review, leaving 305 reviews eligible for AMSTAR rating. AMSTAR scores ranged from 0-10, with a mean of 4.60 (SD 2.18). Mean scores per journal differed significantly ($p=0.008$), with the highest attributed to the Journal of Craniofacial Surgery and the Journal of Plastic, Reconstructive, and Aesthetic Surgery (5.88 and 5.41, respectively). Interrupted time series revealed that introduction of AMSTAR criteria in 2007 did not change average AMSTAR scores. The most commonly met criteria were the presence of a priori inclusion and exclusion criteria (96.1%) and duplicate data extraction (58.0%), while the criteria most often missed were stating conflict of interest for all included studies (2.6%) and inclusion of a list of both included and excluded studies (9.5%).

Discussion:

Significant heterogeneity exists in the quality of systematic reviews published in the plastic surgery literature. Understanding the limitations of the current literature enables future opportunities to increase the quality of systematic reviews guiding clinical decision-making.