

L2

AGREE II: Advancing development, reporting and evaluation of practice guidelines.

Melissa Brouwers¹, Julie Makarski¹, Michelle E. Kho¹, George P. Browman², Jako S. Burgers³, Francoise Cluzeau⁴, Dave Davis⁵, Gene Feder⁶, Beatrice Fervers⁷, Ian D. Graham⁸, Jeremy Grimshaw⁹, Steven E. Hanna¹, Peter Littlejohns¹⁰, Louise Zitzelsberger¹¹

¹McMaster University, Hamilton, Ontario, Canada, ²British Columbia Cancer Agency, Vancouver Island, British Columbia, Canada, ³Dutch Institute for Health Care Improvement CBO, Utrecht, Netherlands, ⁴St. George's Hospital Medical School, London, United Kingdom, ⁵Association of American Medical Colleges, Washington, DC, United States, ⁶University of Bristol, Bristol, United Kingdom, ⁷Centre Leon Berard, Lyon, France, ⁸University of Ottawa, Ottawa, Ontario, Canada, ⁹Ottawa Health Research Institute, Ottawa, Ontario, Canada, ¹⁰National Institute for Health and Clinical Excellence, London, United Kingdom, ¹¹Canadian Partnership Against Cancer, Ottawa, Ontario, Canada

Background: Two studies were conducted to improve the development, reporting, and evaluation of practice guidelines.

Study 1.

Purpose: To conduct an assessment of a new scale, evaluate the performance of the AGREE Instrument and a generic rating scale (GRS), and assess usefulness of and improvements to the instruments.

Methods: 156 participants (clinicians, developers/researchers, policy makers) read a guideline, assessed it with the AGREE and GRS (condition 1) or the GRS alone (condition 2), and completed a series of questionnaires evaluating the usefulness and feasibility of the instrument(s) and required improvements.

Results: All AGREE items and GRS items were rated as useful by participants and *no* differences emerged as a function of user type. All AGREE domains, except editorial independence, predicted users' endorsement of and intentions to use a guideline. The act of applying the AGREE did *not* influence GRS, endorsement or intention to use scores.

Study 2.

Purpose: To test the validity of the draft AGREE II and evaluate the draft User's Manual.

Methods: High and low quality guideline excerpts of draft AGREE II items were randomly assigned to one of two study packages. One quality version of each item was reflected in each package. Thirty participants were randomly assigned to one of the packages. Participants reviewed and rated the guideline content using the draft AGREE II and completed an assessment survey.

Results: Content designed to be of high quality was rated higher than content designed to be of low quality; in 18 of 21 cases, the differences were statistically significant. The Manual was rated by participants as appropriate, easy to use, and able to assist in differentiating good quality guidelines from poor quality guidelines.

Conclusions: The results of these studies led to the final version of the AGREE II, a new international standard in guideline development, reporting, and evaluation.