

Development of a Farsi translation of the AGREE instrument

**and the effects of group discussion on
improving the reliability of the scores**

Dr Arash Rashidian

Dr Reza Yousefi-Nooraie

Knowledge Utilization Research Center
Tehran University of Medical Sciences

Background

- The Appraisal of Guidelines for Research and Evaluation (AGREE)
- Over 20 formal translations of the original AGREE instrument are available
 - also so far six AGREE II translations
- To our knowledge, there is no published report of assessing the reliability and validity of the translated versions

Objectives

- To develop a formal Farsi (Persian) translation of AGREE instrument
 - To aid health care professionals and decision makers to appraise clinical guidelines
- To assess reliability of the translated tool
- To assess the effects of group discussion on the reliability of the AGREE scores

Methods - translation

- *Multi-step process of translation:*
- Two independent translations
 - Health service researchers
- Independent check and comparison of translations / developing the third version
 - Experienced guideline developer and methodologist

Methods - translation

- Review by three methodologists and three guideline users / developing pre-final version
 - rigor of translation, face validity and fluency, consistency of terminology
- Distribution to all heads of medical research centers in Iran
 - received few minor comments at this stage

نقد راهنمای بالینی در پژوهش و ارزشیابی

مجله پژوهش‌های علمی و ارزشیابی

AGREE
INSTRUMENT
Validated Form (Persian) translation

The AGREE Collaboration
September 2001

ترجمه از دیومنت ۲۰۰۱

دائرة طب و تجهیزات آموزشی وزارت بهداشت، درمان و آموزش پزشکی
معاونت سلامت

مرکز تحقیق و توسعه سیاست‌های آموزشی در نظام سلامت
مؤسسه علوم پزشکی تهران

Center for Academic and Health Policy (CAHP)
Tehran University of Medical Sciences

Methods - reliability

- Five educational workshops on guideline development and appraisal using AGREE
- 11 to 30 participants in each workshop
 - Physicians, nurses, methodologists, etc
- Using AGREE to appraise 11 selected guidelines
 - All guidelines in English
 - Three clinical domains: neonatology, hepatitis and dyspepsia, thyroid nodules
 - Topics in line with participants interests

Methods - reliability

- *Appraisal process:*
- Participants divided in small groups comprising of 3 to 5 people
 - Multi-disciplinary groups
 - Each group received one guideline, randomly assigned
- The AGREE English and Farsi versions were distributed randomly
- 96 appraisals in total

Methods - reliability

- Independent appraisal of the guidelines
- Small groups discuss the scores with each other
- Opportunity to revise scores after discussions

Methods - analysis

- Comparing standardized mean (SDs) domain scores
 - Farsi vs English tool; pre vs post discussion
- General linear model with the domain scores as dependent variable
 - AGREE language and time of assessment (session) as fixed factors
 - Guideline names as random factor
- Domain specific ICCs

Guideline title	Guideline development organization	Country	Date	Appraisers per instrument (English, Farsi)
Guidelines for good practice in the management of neonatal respiratory distress syndrome	Second working group of the British Association of Perinatal Medicine	UK	1998	11(4,7)
Recommendations for neonatal surfactant therapy	Canadian Pediatric Society	Canada	2005	12(5,7)
Dyspepsia: A national clinical guideline	Scottish Intercollegiate Guidelines Network (SIGN)	Scotland	2003	8 (4,4)
Management of hepatitis C infection: regional guideline	Iranian Hepatitis Group	Iran	2004	9(4,5)
Management guidelines for patients with thyroid nodules and differentiated thyroid cancer	The American Thyroid Association Guidelines Taskforce	USA	2006	8(4,4)
Clinical practice guidelines in oncology: thyroid carcinoma v.2	National Comprehensive Cancer Network	USA	2007	8(4,4)
Medical/surgical guidelines for clinical practice: management of thyroid carcinoma	American Association of Clinical Endocrinologists and Associazione Medici Endocrinologi	USA	2001	8(4,4)
Medical guidelines for clinical practice for the diagnosis and management of thyroid nodules	American Association of Clinical Endocrinologists and Associazione Medici Endocrinologi	USA	2006	8(4,4)
I-131 remnant ablation after thyroidectomy for papillary or follicular thyroid cancer	Cancer Care	Canada	2005	8(4,4)
Guidelines for the management of thyroid cancer	British Thyroid Association, of Physicians	UK	2007	8(4,4)
Procedure guideline for therapy of thyroid disease with 131Iodine	The Society of Nuclear Medicine	USA	2002	8(4,4) ¹¹

ICC values for the reliability of domain scores

Domain	Before discussion			After discussion		
	English	Farsi	Error variance	English	Farsi	Error variance
Scope and purpose	27.4	0	686.6	34.3	10	503.7
Stakeholder involvement	49.9	38.5	171.7	63.2	49.2	117.0
Rigor of development	64.3	42.7	186.6	82.7	64.6	89.8
Clarity and presentation	42.8	35.2	229.0	57.9	48.4	117.7
Applicability	14.1	18.7	286.4	23.8	23.6	219.2
Editorial independence	38.7	33.3	453.1	52.8	64	315.9

Similar observations for mean differences between scores

Reliability of Farsi (Persian) instrument

- Average domain scores obtained from the two versions of the instrument were similar
- The translated tool is sufficiently reliable
- Still the Farsi version showed, non significant, lower reliability
 - Why?
- Technical terms do not have widely shared counterpart translations in Farsi
 - might have been interpreted differently by different users

Reliability of AGREE

- Guideline appraisal tools like AGREE showed non-optimal reliability measures in previous studies
 - Need averaging the assessments of at least four raters to increase reliability
- Revisions in AGREE-II to improve reliability
 - Item wordings
 - Domain structures
 - 7-point scale instead of 4-point scale

Reliability of AGREE

- Modifications did not satisfactorily improve reliability
- Further work still required
- Including changes in appraisal strategies
- More studies on appraisal process are required

Reliability of AGREE

- We suggest a two-step rating by AGREE instrument, pre and post group discussion
- Using group discussion to facilitate consensus and reduce misunderstandings and errors
- Especially where appraisers are not professional appraisers and have limited time
 - happens frequently in clinical practice
 - Revisions should be done individually to reduce the risk of dominance of powerful members

Further suggestions

Item	Current domain	Suggested alternative domain
6. The target users of the guideline are clearly defined	Stakeholder involvement	Scope and purpose
7. The guideline has been piloted among target users	Stakeholder involvement	Applicability*
18. The guideline is supported with tools for application	Clarity and presentation	Applicability*

* Similar changes to what proposed by the participants are incorporated in the AGREE II

■ Study to be published in JECP – *in press*

Journal of **Evaluation in Clinical Practice**
International Journal of Public Health Policy and Health Services Research

Journal of Evaluation in Clinical Practice ISSN 13

Development of a Farsi translation of the AGREE instrument, and the effects of group discussion on improving the reliability of the scores

Arash Rashidian MD PhD¹ and Reza Yousefi-Nooraie MD²

¹Associate Professor, Department of Health Management and Economics, School of Public Health, Tehran University of Medical Sciences, Iran and Deputy Director for Research, Knowledge Utilization Research Center, Tehran University of Medical Sciences, Tehran, Iran

²Researcher, Knowledge Utilization Research Center, Tehran University of Medical Sciences, Tehran, Iran and PhD Student, Department of Epidemiology and Biostatistics, McMaster University, Hamilton, Canada

Thank you