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
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)

### ICC values for the reliability of domain scores

<table>
<thead>
<tr>
<th>Domain</th>
<th>Before discussion</th>
<th>After discussion</th>
<th>Error variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>Farsi</td>
<td>English</td>
</tr>
<tr>
<td>Scope and purpose</td>
<td>27.4</td>
<td>0</td>
<td>686.6</td>
</tr>
<tr>
<td>Stakeholder involvement</td>
<td>49.9</td>
<td>38.5</td>
<td>171.7</td>
</tr>
<tr>
<td>Rigor of development</td>
<td>64.3</td>
<td>42.7</td>
<td>186.6</td>
</tr>
<tr>
<td>Clarity and presentation</td>
<td>42.8</td>
<td>35.2</td>
<td>229.0</td>
</tr>
<tr>
<td>Applicability</td>
<td>14.1</td>
<td>18.7</td>
<td>286.4</td>
</tr>
<tr>
<td>Editorial independence</td>
<td>38.7</td>
<td>33.3</td>
<td>453.1</td>
</tr>
</tbody>
</table>

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

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

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

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

---

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

*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 PhD and Reza Yousefi-Nooria MD

---

Thank you