Identification of validated questionnaires to measure adherence to pharmacological antihypertensive treatments

Beatriz Pérez-Escamilla¹
Lucía Franco-Trigo¹
Joanna C Moullin²
Fernando Martínez-Martínez¹
José P García-Corpaś
¹ Academic Centre in Pharmaceutical Care, Faculty of Pharmacy, University of Granada, Granada, Spain; ² Graduate School of Health, Faculty of Pharmacy, University of Technology Sydney, Sydney, NSW, Australia

Background: Low adherence to pharmacological treatments is one of the factors associated with poor blood pressure control. Questionnaires are an indirect measurement method that is both economic and easy to use. However, questionnaires should meet specific criteria, to minimize error and ensure reproducibility of results. Numerous studies have been conducted to design questionnaires that quantify adherence to pharmacological antihypertensive treatments. Nevertheless, it is unknown whether questionnaires fulfill the minimum requirements of validity and reliability. The aim of this study was to compile validated questionnaires measuring adherence to pharmacological antihypertensive treatments that had at least one measure of validity and one measure of reliability.

Methods: A literature search was undertaken in PubMed, the Excerpta Medica Database (EMBASE), and the Latin American and Caribbean Health Sciences Literature database (Literatura Latino-Americana e do Caribe em Ciências da Saúde [LILACS]). References from included articles were hand-searched. The included papers were all that were published in English, French, Portuguese, and Spanish from the beginning of the database’s indexing until July 8, 2013, where a validation of a questionnaire (at least one demonstration of the validity and at least one of reliability) was performed to measure adherence to antihypertensive pharmacological treatments.

Results: A total of 234 potential papers were identified in the electronic database search; of these, 12 met the eligibility criteria. Within these 12 papers, six questionnaires were validated: the Morisky–Green–Levine; Brief Medication Questionnaire; Hill-Bone Compliance to High Blood Pressure Therapy Scale; Morisky Medication Adherence Scale; Treatment Adherence Questionnaire for Patients with Hypertension (TAQPH); and Martín–Bayarre–Grau. Questionnaire length ranged from four to 28 items. Internal consistency, assessed by Cronbach’s α, varied from 0.43 to 0.889. Additional statistical techniques utilized to assess the psychometric properties of the questionnaires varied greatly across studies.

Conclusion: At this stage, none of the six questionnaires included could be considered a gold standard. However, this revision will assist health professionals in the selection of the most appropriate tool for their individual circumstances.

Keywords: validation, hypertension, medication, compliance, scale, validity, reliability

Introduction

Hypertension is a major public health concern worldwide. It is one of the leading risk factors for cardiovascular disease and is associated with a decrease in a patient’s quality of life and an increase in the probability of health complications. An estimated 14% of all deaths worldwide are attributable to high blood pressure (systolic blood pressure >140 mmHg; diastolic blood pressure >90 mmHg). In Spain, 54% of...
cardiovascular-related deaths in people older than 50 years are due to hypertension.\(^3\)

It is widely accepted that there are relationships between poor blood pressure control and a patient’s lack of adherence to antihypertensive treatment, as well as lack of effectiveness to antihypertensive treatments.\(^4,5\) Research has been conducted in an attempt to reduce nonadherence.\(^6,7\) In spite of this, a study showed the mean prevalence of nonadherence to antihypertensive treatments to be 30%.\(^8\)

Several terms are found in biomedical literature to explain the degree to which patients follow prescribed medication directions, including the terms “adherence” and “compliance”. Haynes et al defined “compliance” in 1979 as the extent to which a person’s behavior (in terms of taking medications, following diets or executing lifestyle changes) coincides with medical care or health advice; thus, noncompliance is the extent to which these instructions are not accomplished.\(^9\) The World Health Organization (WHO) combined the definitions developed by Haynes et al with one by Rand\(^10\) to obtain their definition of adherence to long-term treatment as

the extent to which a person’s behavior (in terms of taking medications, following diets or executing lifestyle changes) corresponds to the recommendations agreed by a health care provider.\(^11\)

Later, Osterberg and Blaschke stated the term “adherence”

… is preferred by many health care providers, because compliance suggests that the patient is passively following the doctor’s orders and that the treatment plan is not based on a therapeutic alliance or contract established between the patient and the physician.\(^12\)

Regardless of the definition used, health discipline, or health problem, measuring adherence has been complicated. Both direct (biological measures) and indirect (pill counts, patient-kept diaries of medication-taking, and questionnaires) methods have been used.\(^13,14\) Biological measures are considered a gold standard due to their objectivity, but their high cost is prohibitive. On the other hand, pill counts are simple, but they do not guarantee patient collaboration, and as such, results of pill counts may be inaccurate. Questionnaires are an alternative. Even though questionnaires have the disadvantage of overestimating patient adherence or nonadherence,\(^15\) their advantage is that they are easy to use and are relatively inexpensive.\(^16\) Moreover, questionnaires are the most common method used in clinical settings because they have the ability to provide information about a patient’s reasons for not adhering to prescribed treatments.\(^17\) A number of scales of adherence to antihypertensive medication have been developed, but a questionnaire considered to be a gold standard does not exist. The most widely used\(^17-20\) questionnaire is that designed by Morisky–Green–Levine (MGL)\(^21\) in 1986. This is a unidimensional questionnaire containing four items.

In order to be a useful tool, a questionnaire must be valid and reliable.\(^22\) Scale validity refers to

the degree of confidence we can have that the measurement corresponds to the reality of the phenomenon that is being measured,\(^23\) that is,

the degree to which an instrument measures what it is supposed to measure.\(^24\)

Whereas reliability is

the extent to which a measure yields the same number or score each time it is administered when the construct being measured has not changed.\(^25\)

A number of questionnaires used in daily clinical practice do not reach the minimum standards for validity and reliability.\(^26\) Given this, it is necessary to compile an exhaustive list of the available validated tools measuring adherence to pharmacological antihypertensive treatment.

The aim of this work was to compile the questionnaires designed to measure adherence to pharmacological antihypertensive treatment that have at least one test of validity and one test of reliability.

**Materials and methods**

A literature search was undertaken in three electronic databases: US National Library of Medicine (PubMed), Excerpta Medica Database (EMBASE), and the Latin American and Caribbean Health Sciences Literature database (Literatura Latino-Americana e do Caribe em Ciências da Saúde [LILACS]). The search terms were: “patient adherence”, “patient compliance [Medical Subject Headings {MeSH}]]”, “compliance”, “predictive validity”, “content validity”, “concurrent validity”, “convergent validity”, “discriminant validity”, “construct validity”, “psychometric properties”, “clinimetric properties”, “test-retest reliability”, “temporal stability”, “interobserver agreement”, “internal homogeneity”, “internal consistency”, “questionnaires [MeSH]”, “reproducibility of results [MeSH]”, “cronbach’s alpha”, and “hypertension [MeSH]”. Keywords were truncated in the LILACS database to avoid the loss of papers that could be of interest. In addition, the search was complemented by reference lists from the included papers.
Inclusion criteria
Articles needed to include the validation of a questionnaire to measure patient adherence to pharmacological antihypertensive treatments, published from the beginning of the database’s indexing until July 8, 2013. The articles had to include at least one validity test (content-, construct-, or criterion-related) and one reliability test (stability, equivalence, or homogeneity) of the questionnaire. Language limits were English, Spanish, French, and Portuguese.

Article selection
Duplications were removed. Two independent authors read the titles and abstracts to select the articles that met the inclusion criteria. If doubts appeared, the whole article was read. If there was disagreement, a third author arbitrated the debate between the two first authors.

Data extraction
Two authors extracted data independently. These data referred to the characteristics of the studied population and to the psychometric properties of the questionnaire being tested.26 Afterwards, extracted data was checked for any disagreement. When there was disagreement, the third author resolved the conflict.

Results
A total 234 articles were retrieved. Three were not available. From the remaining 231, 52 were duplicated. Of these, 104 articles were removed by title, 46 by abstract, and finally, 17 were removed after reading the whole article. Ultimately, 12 articles were included in the revision (Figure 1).

Articles were removed if they did not meet the inclusion criteria, that is, they did not validate a questionnaire for measuring adherence to antihypertensive treatments or demonstrate the required psychometric properties. For example,

Figure 1 Flow of information through the different phases of the systematic review.

Abbreviations: EMBASE, Excerpta Medica Database; LILACS, Literatura Latino Americana e do Caribe em Ciências da Saúde (Latin American and Caribbean Health Sciences Literature).
some articles did not measure treatment adherence but rather, reported the patients' understanding about hypertension treatment, while others measured quality of life in hypertensive patients, patient knowledge about the disease, or patient satisfaction with their medication. Furthermore, there were articles that measured adherence in diseases other than hypertension and in other fields, such as nutrition. Articles in which adherence was the subject but measured related factors (aptitudes, beliefs, etc) instead of classifying patients as adherent or nonadherent, were similarly eliminated.

The 12 articles selected included 15 validation processes for six questionnaires that measured adherence to pharmacological antihypertensive treatments (Table 1). The questionnaires were the following:

- MGL (two validations)\(^{21,27}\)
- Brief Medication Questionnaire (BMQ) (one validation)\(^{33}\)
- Hill-Bone Compliance to High Blood Pressure Therapy Scale (HB Comp Scale) (six validations)\(^{28–30}\)
- Morisky Medication Adherence Scale (MMAS-8) (three validations)\(^{31–33}\)
- Treatment Adherence Questionnaire for Patients with Hypertension (TAQPH) (one validation)\(^{34}\)
- Martín–Bayarre–Grau (MBG) (two validations)\(^{35,36}\)

### Content validity

Content validity was assessed in four questionnaires: HB Comp Scale\(^{28}\), TAQPH, MMAS-8 (Urdu version),\(^{31}\) and MBG (validated in Cuba).\(^{35}\) In all four validation studies, the authors used an expert panel. In addition, for item generation, two questionnaires (HB Comp Scale and TAQPH) included a narrative review, and one (TAQPH) performed a focus group.

### Construct validity

Construct validity was assessed in eleven of 15 validation processes (Table 2). The Urdu version of the MMAS-8\(^{33}\) evaluated both convergent validity, by comparing the results to the MGL,\(^{21}\) and known group validity, comparing the results of the questionnaire to blood pressure control. The remaining construct validity assessments were performed by factor analysis. In the MGL,\(^{21}\) HB Comp Scale,\(^{28}\) Hill-Bone Medication Adherence-Korean version (HBMA-K) scale (a validation of HB Comp Scale),\(^{30}\) and MMAS-8 (English\(^{32}\) and French\(^{31}\) versions) the factor analysis yielded a one-dimensional solution. On the other hand, the TAQPH\(^{34}\) questionnaire yielded six factors, and the Colombian MBG\(^{36}\) yielded five.

Total variance ranged from 68.7%, in the MBG validated in Cuba,\(^{35}\) to 27%, in the HB Comp Scale validated in young African American males.\(^{28}\)

### Table 1

<table>
<thead>
<tr>
<th>Author, year of publication</th>
<th>Questionnaire</th>
<th>Mean age (SD)</th>
<th>Sex % female, (n)</th>
<th>Language</th>
<th>Setting</th>
<th>Sample (n)</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morisky et al(^{31}) 1986</td>
<td>MGL</td>
<td>54(^{a})</td>
<td>70.00 (–)</td>
<td>English</td>
<td>Hospital (outpatient clinics)</td>
<td>290</td>
<td>4</td>
</tr>
<tr>
<td>Ben et al(^{27}) 2012</td>
<td>BMQ vs MGL</td>
<td>66.60 (13.20)</td>
<td>64.60 (133)</td>
<td>Portuguese</td>
<td>Primary health care unit</td>
<td>206</td>
<td>11</td>
</tr>
<tr>
<td>Kim et al(^{30}) 2000</td>
<td>HB Comp Scale</td>
<td>41.30 (5.30)</td>
<td>0.00 (0)</td>
<td>Portuguese</td>
<td>Outpatient clinic</td>
<td>Study 1: 139</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>59.20 (13.10)</td>
<td>69.20 (236)</td>
<td>English</td>
<td></td>
<td>Study 2: 341</td>
<td>14</td>
</tr>
<tr>
<td>Krousel-Wood et al(^{35}) 2005</td>
<td>HB Comp Scale</td>
<td>69.00 (–)</td>
<td>49.00 (121)</td>
<td>English</td>
<td>Elderly community-dwelling</td>
<td>239</td>
<td>14</td>
</tr>
<tr>
<td>Lambert et al(^{36}) 2006</td>
<td>Xhosa HB Comp Scale</td>
<td>52.00 (7.60)</td>
<td>51.00 (50)</td>
<td>Xhosa</td>
<td>Community health centers</td>
<td>79</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Xhosa Adapted HB Comp Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>82</td>
<td>10</td>
</tr>
<tr>
<td>Song et al(^{31}) 2011</td>
<td>HBMA-K</td>
<td>52.54 (5.40)</td>
<td>56.80 (88)</td>
<td>Korean</td>
<td>Baltimore-Washington metropolitan area</td>
<td>Study 1: 155</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70.88 (5.40)</td>
<td>70.50 (261)</td>
<td></td>
<td></td>
<td>Study 2: 370</td>
<td></td>
</tr>
<tr>
<td>Morisky et al(^{31}) 2008</td>
<td>MMAS-8</td>
<td>52.50 (12.20)</td>
<td>59.20 (809)</td>
<td>English</td>
<td>Hospital</td>
<td>1,367</td>
<td>8</td>
</tr>
<tr>
<td>Korb-Savoldelli et al(^{31}) 2012</td>
<td>MMAS-8</td>
<td>55.70 (14.60)</td>
<td>42.70 (85)</td>
<td>French</td>
<td>Hospital, day care unit</td>
<td>199</td>
<td>8</td>
</tr>
<tr>
<td>Saleem et al(^{31}) 2012</td>
<td>MMAS-8</td>
<td>39.50 (6.93)</td>
<td>28.20 (31)</td>
<td>Urdu</td>
<td>Hospital</td>
<td>110</td>
<td>8</td>
</tr>
<tr>
<td>Ma et al(^{31}) 2012</td>
<td>TAQPH</td>
<td>59.68 (10.69)</td>
<td>58.30 (162)</td>
<td>Chinese</td>
<td>Hospital</td>
<td>278</td>
<td>28</td>
</tr>
<tr>
<td>Martín Alfonso et al(^{31}) 2008</td>
<td>MBG</td>
<td>–</td>
<td>–</td>
<td>Spanish</td>
<td>Polyclinic</td>
<td>114</td>
<td>12</td>
</tr>
<tr>
<td>Martínez et al(^{36}) 2011</td>
<td>MBG</td>
<td>Not indicated(^{a})</td>
<td>81.00 (115)</td>
<td>Spanish</td>
<td>Health care center</td>
<td>142</td>
<td>12</td>
</tr>
</tbody>
</table>

Notes: \(^{a}\)Median. \(^{b}\)Range 37–90 years.

Abbreviations: BMQ, Brief Medication Questionnaire; HB Comp Scale, Hill-Bone Compliance to High Blood Pressure Therapy Scale; HBMA-K, Hill-Bone Medication Adherence-Korean version scale; MBG, Martín–Bayarre–Grau; MGL, Morisky–Green–Levine; MMAS-8, Morisky Medication Adherence Scale; SD, standard deviation; TAQPH, Treatment Adherence Questionnaire for Patients with Hypertension.
Table 2 Validity and reliability assessment

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Construct validity</th>
<th>Criterion validity</th>
<th>Stability (test–retest)</th>
<th>Homogeneity (internal consistency)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Item-total correlations</td>
</tr>
<tr>
<td>MGL&lt;sup&gt;21&lt;/sup&gt;</td>
<td>PCFA: one factor</td>
<td>Comparative test: BP control.</td>
<td>–</td>
<td>0.479–0.561</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Concurrent validity: r=0.43 (6 months)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Predictive validity: r=0.58 (42 months)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Student’s t-test = 6.43, P &lt; 0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Medication intake vs BP control: $R^2=0.33$, $P&lt;0.01$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$S=81.00%$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$SP=44.00%$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$PPV=75.00%$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$NPV=52.00%$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMQ&lt;sup&gt;22&lt;/sup&gt;</td>
<td>–</td>
<td>–</td>
<td>BMQ: r=0.83, P &lt; 0.001</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Screen 1: r=0.84, P=0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Screen 2: r=0.86, P=0.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Screen 3: r=0.94, P=0.120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGL&lt;sup&gt;27&lt;/sup&gt;</td>
<td>–</td>
<td>–</td>
<td>MGL: r=0.70, P=0.020</td>
<td>–</td>
</tr>
<tr>
<td>HB Comp Scale&lt;sup&gt;23&lt;/sup&gt;</td>
<td>Factor analysis variance</td>
<td>Comparative test: BP control.</td>
<td>–</td>
<td>Range: -0.02 to 0.60</td>
</tr>
<tr>
<td>Study 1</td>
<td>(%) 27.00, one factor</td>
<td>Predictive validity: r=0.21, P=0.030</td>
<td></td>
<td>Mean: 0.34 (SD 0.17)</td>
</tr>
<tr>
<td>Study 2</td>
<td>Factor analysis variance</td>
<td>Comparative test: BP control.</td>
<td>–</td>
<td>Range: 0.01–0.64</td>
</tr>
<tr>
<td></td>
<td>(%) 35.00, one factor</td>
<td>Predictive validity: r=0.16, P&lt;0.001</td>
<td></td>
<td>Mean: 0.46 (SD 0.16)</td>
</tr>
<tr>
<td>HB Comp Scale&lt;sup&gt;7&lt;/sup&gt;</td>
<td>Factor analysis</td>
<td>Not available&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Range: 0.13–0.59&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Full questionnaire: 0.43</td>
</tr>
<tr>
<td>(elderly)</td>
<td>two factors</td>
<td></td>
<td></td>
<td>– Medication compliance subscale: 0.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Salt intake subscale: 0.49</td>
</tr>
<tr>
<td>Xhosa HB Comp Scale&lt;sup&gt;29&lt;/sup&gt;</td>
<td>–</td>
<td>Comparative test: BP control.</td>
<td>–</td>
<td>Range: 0.10–0.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Predictive validity: questionnaire vs</td>
<td></td>
<td>Mean: 0.38 (SD: 0.18)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Mean SBP: $p=0.17$, $p=0.140$</td>
<td></td>
<td>Subscale I (salt intake): 0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Mean DBP: $p=0.18$, $p=0.130$</td>
<td></td>
<td>Subscale II (appointment-keeping): no result&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Xhosa Adapted HB Comp Scale&lt;sup&gt;29&lt;/sup&gt;</td>
<td>–</td>
<td>Comparative test: BP control.</td>
<td>–</td>
<td>Range: 0.31–0.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Predictive validity: questionnaire vs</td>
<td></td>
<td>Mean: 0.45 (SD 0.11)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Mean SBP: $p=0.19$, $p=0.080$</td>
<td></td>
<td>Subscale III (medication compliance): 0.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Mean DBP: $p=0.21$, $p=0.050$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
### Table 2 (Continued)

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Construct validity</th>
<th>Criterion validity</th>
<th>Stability (test–retest)</th>
<th>Homogeneity (internal consistency)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Item-total correlations</td>
</tr>
<tr>
<td><strong>HBMA-K</strong>&lt;sup&gt;30&lt;/sup&gt;</td>
<td>Factor analysis variance (%): 35.40, one factor</td>
<td>Comparative test: questionnaire vs – HBP knowledge with 26 item test – Beliefs about HBP – BP control</td>
<td>–</td>
<td>Range: 0.37–0.63</td>
</tr>
<tr>
<td><strong>MMAS-8</strong>&lt;sup&gt;31&lt;/sup&gt; (English)</td>
<td>Factor analysis CFA: one factor</td>
<td>Concurrent validity: questionnaire vs MGL r = 0.64, P &lt; 0.05 S = 92.00% SP = 53.00% Predictive validity: questionnaire vs – BP control (χ² = 6.6, P &lt; 0.05) – knowledge, attitude, social support, stress coping, medication complexity, and patient satisfaction with clinic visits&lt;sup&gt;14,42&lt;/sup&gt; (P &lt; 0.05 for all, except for attitude)</td>
<td>–</td>
<td>Range: 0.3038–0.5896</td>
</tr>
<tr>
<td><strong>MMAS-8</strong>&lt;sup&gt;31&lt;/sup&gt; (French)</td>
<td>Factor analysis PCFA variance (%): 27.50, one factor CFA: one factor</td>
<td>–</td>
<td>ICC = 0.68 (95% CI; 0.63–0.72)</td>
<td>Range: −0.05 to 0.43</td>
</tr>
<tr>
<td><strong>MMAS-8</strong>&lt;sup&gt;31&lt;/sup&gt; (Urdu)</td>
<td>Convergent validity, comparative test: questionnaire vs MGL r = 0.765, P &lt; 0.001 Known group validity vs BP control χ² = 19.996, P &lt; 0.001</td>
<td>–</td>
<td>ICC meaningful (P &lt; 0.001)</td>
<td>Range: 0.30–0.48</td>
</tr>
<tr>
<td><strong>TAQPH</strong>&lt;sup&gt;34&lt;/sup&gt;</td>
<td>EFA variance (%): 62.50, six factors CFA: six factors</td>
<td>Concurrent validity, comparative test: questionnaire vs MMAS-8; GSES&lt;sup&gt;43&lt;/sup&gt; MMAS-8 → r = 0.76, P &lt; 0.01 GSES → r = 0.69, P &lt; 0.01</td>
<td>ICC = 0.82</td>
<td>Range: 0.31–0.66</td>
</tr>
<tr>
<td><strong>MBG</strong>&lt;sup&gt;35&lt;/sup&gt;</td>
<td>PCFA variance (%): 68.70, three factors</td>
<td>–</td>
<td>–</td>
<td>Range: 0.28–0.70</td>
</tr>
<tr>
<td><strong>MBG</strong>&lt;sup&gt;35&lt;/sup&gt;</td>
<td>PCFA variance (%): 63.40, five factors</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

**Notes:**<sup>a</sup> Longitudinal data to assess test–retest reliability were not available. <sup>b</sup>The appointment-keeping subscale of the scale consists of only two items and was not included in this analysis. <sup>c</sup>The appointment-keeping subscale did not provide an interpretable result, probably as a result of the South African public health system, in which reappointments are standard procedure.

**Abbreviations:** BMQ, Brief Medication Questionnaire; BP, blood pressure; CFA, confirmatory factorial analysis; CI, confidence interval; DBP, diastolic blood pressure; EFA, exploratory factorial analysis; GSES, General Self-Efficacy Scale; HB Comp Scale, Hill-Bone Compliance to High Blood Pressure Therapy Scale; HBMA-K, Hill-Bone Medication Adherence – Korean version (scale); HBP, high blood pressure; ICC, intraclass correlation coefficient; MBG, Martín–Bayarri–Grau (scale); MGL, Morisky–Green–Levine (scale); MMAS-8, Morisky Medication Adherence Scale; NPV, negative predictive value; PCFA, principal components factorial analysis; PPV, positive predictive value; S, sensitivity; SBP, systolic blood pressure; SD, standard deviation; SP, specificity; TAQPH, Treatment Adherence Questionnaire for Patients with Hypertension.
Criterion validity

Four validations (MGL, HBMA-K, MMAS-8 (English), and TAQPH) assessed concurrent validity, and six evaluated predictive validity (MGL, Xhosa HB Comp Scale, HB Comp Scale [two validations], and the MMAS-8 [English]). The comparison tests are detailed in Table 2.

Reliability

Cronbach’s $\alpha$ was used to estimate internal consistency as a reliability measure in every validation. The MBG validated in Cuba obtained the highest coefficient ($\alpha=0.889$), while the lowest ($\alpha=0.43$) was obtained in the HB Comp Scale validated in an elderly population, published by Krousel-Wood et al. Item-total scale correlation was also used in all but three validations (Table 2).

The MMAS scale in Urdu and French, the TAQPH, and the BMQ evaluated temporal stability with the test–retest method. In each of these validations, temporal stability was demonstrated with a moderate to high correlation, except for the TAQPH questionnaire, which had an intraclass correlation coefficient of 0.82.

Discussion

This review provides health professionals with a report summarizing the evidence across different questionnaires according to the psychometric properties evaluated. The review did not intend to determine or indicate an optimum questionnaire but rather, to compile validated scales that measure antihypertensive treatment adherence. Consequently, health professionals can choose the most appropriate questionnaire, depending on their circumstances. To our knowledge there is no similar published study.

A large number of authors have designed questionnaires in order to measure adherence to antihypertensive treatments. Nevertheless, no questionnaire could be considered a gold standard. Any tool to be used as a measure must demonstrate validity and reliability, including questionnaires. Furthermore, validity and reliability must be measured in each sample or have been conducted in a comparable sample since it is not possible to extrapolate results among different populations. Differences in validity and reliability results obtained across versions of the MMAS-8 can be seen as an example.

A sensitive search strategy was designed to ensure the collection of all papers validating questionnaires to measure adherence to pharmacological antihypertensive treatment. Inclusion criteria in this review required that articles have at least one validity test and one of reliability. Taking into account these criteria, such well-known questionnaires as the BMQ (English version), the Haynes–Sackett and Batalla questionnaires, etc were not included.

The validity of a questionnaire (the confirmation that it measures what it is supposed to measure) is evaluated in several ways, as: content validity, construct validity, and criterion validity.

The methods most commonly used to assess content validity are expert opinion and systematic review. However, researchers often choose only one of these options. From the 15 validations found, only four studied content validity, and even these did not explain in detail the process followed to assess it. It is suspected that in the remaining questionnaires, authors assumed, without testing, content validity was present. In the four validations that assessed content validity, the technique utilized was an expert panel. None of the authors based content validity on a systematic review. This method could be of great interest, but it can increase costs and research time.

The most common technique to assess construct validity was factor analysis. A one-dimensional solution emerged in the majority of the scales, but the factor explained only a small percentage of the total response variability. As an example, the HB Comp Scale explained up to 35% of the total variance. Consequently questionnaires should consider including further dimensions and items to explain more of the total response variance. This could be observed in the TAQPH, which had 28 items and six factors, and was able to explain 62.5% of the variance. The total explained variance is considered an indicator of how factors extracted from a questionnaire actually correspond to patient answers. When little variance is explained, the scale shows a deficiency that indicates lack of variables or a poor initial theoretical construct. However, it is also important that a questionnaire does not have too many items, to minimize patient resistance to being interviewed and interview duration. As such, the number of items needs to be balanced between explaining variance and user acceptance.

Construct validity assessment by convergent validity was completed in two questionnaires. The correlation between the Urdu version of the MMAS-8 and the same language version of the MGL was checked. A high correlation was found between the scales ($r=0.765, P<0.001$), which in theory confirmed that both questionnaires measured the same construct. However such a correlation is not surprising as the MMAS-8 was based on the MGL four-item questionnaire, onto which more questions were added.
Criterion validity is traditionally defined as
the correlation of a scale with some other measure of the
trait or disorder under study, ideally a gold standard which
has been used and accepted in the field.46

Criterion validity consists of two types: concurrent and
predictive. In concurrent validity, both the new scale and
a criterion measure are applied simultaneously and subse-
quently correlated. On the other hand, when using predictive
validity, the criterion measure will not be available until
sometime in the future. This implies that testing predictive
validity may extend the time of study.

With the exception of six validations, the remainder tested
criterion validity. The validation of HBMA-K10 obtained a
low correlation to systolic blood pressure control (r=0.18;
PR0.01) and diastolic blood pressure (r=0.24; PR0.01).
This low correlation could be due to the fact that medication
adherence is not perfectly correlated to blood pressure control.
That is, it is possible for a patient to have high medication
adherence and to not have controlled blood pressure due to
another factor, such as lack of effectiveness of the treatment.
Nevertheless, it is assumed that there is a direct relationship
between adherence and blood pressure control.47 On the other
hand, the MMAS-8,32 compared with the MGL21 to assess
concurrent validity, obtained a moderate to high correlation
(r=0.64; PR0.05). It is necessary to keep in mind that the
development of the MMAS-8 was based on the MGL, as
stated previously.

In addition to verifying the validity of a questionnaire, it
is necessary to check the results are reproducible. That is, it
is necessary to assess its reliability. From a technical point
of view, it could be said that reliability is

a ratio of the variability between individuals to the total
variability in the scores; in other words, the reliability is a
measure of the proportion of the variability in scores which
was due to true differences between individuals.46

One of the methods used to measure reliability is internal
consistency. Internal consistency describes reliability estima-
tions based on the average correlation among items in a test.46
A Cronbach’s α value is considered acceptable when above
0.70.43 Only two validations obtained Cronbach’s α values
well below 0.70 (the HB Comp Scale [elderly] [0.43] and
MMAS-8 [French]11 [0.54]), indicating the items were not
highly related. In the remaining questionnaires, Cronbach’s α
varied from 0.61 (in the MGL)31 to 0.889 (in the MGB).35
However, it should be noted that Cronbach’s α is sensitive
to the number of items and the sample’s variance.

Cronbach’s α is assessed from a single administra-
tion of the test, not taking into account variations in time
or between administrators, and as such, it can provide an
optimistic interpretation of the reliability of a questionnaire.
For this reason, it is advisable that internal consistency
is accompanied by a stability measure (test–retest) or by
interobserver equivalence.46 In the present study, six validi-
tions provided stability data and none checked interobserver
verifications.

As previously stated, the intent of this article was to
gather all questionnaires measuring patient adherence to
antihypertensive treatments that contain at least one validity
measure and one reliability measure. Furthermore the
authors do not consider any of the questionnaires to have
sufficient reliability or validity to be highly recommended.
Nevertheless, if one had to choose a questionnaire, the
authors offer these considerations.

For research, the TAQPH may be advisable. The
TAQPH contains the greatest number of items, potentially
providing more information to the research investigation,
explains a high percentage of variance, and has sound
indicators for both stability and criterion validity. The
MBG also contains a large number of items but cannot be
endorsed as it lacks criterion validity, stability, and item-
total correlations.

In clinical practice it is necessary that a questionnaire
is acceptable to both health care professionals and patients
(containing the fewest possible items), is quick and easy to
use, and shows good sensitivity, specificity, positive predic-
tive value, and negative predictive value. The MGL, BMQ,
HB Comp Scale, and MMAS-8 have similar indicators of
reliability and validity; however, only the MMAS-8 and
MGL show the probability of a clinician deciding that a non-
adherent patient is not controlled (positive predictive value)
or that an adherent patient is controlled (negative predictive
value). As such, in clinical practice, both the MMAS-8 and
MGL may be recommended. However as previously noted,
the MMAS-8 was developed by Morisky et al in 200832 based
on the MGL and as such, has an improved capacity to collect
information (by having four additional items), sensitivity,
specificity, positive predictive value, and negative predictive
value, while maintaining acceptable validity and reliability.
Consequently in clinical practice, the MMAS-8 may be an
appealing option.

Limitations
This study could be influenced by publication bias as a conse-
quence of the general tendency to publish only positive results.
Language bias could also have appeared. In spite of including studies published in four languages, it is possible that some studies were not included because of being published in other languages. Moreover, despite the strict methodology followed by reviewers, it is possible that selection bias appeared as a consequence of the lack of availability of some articles.

**Conclusion**

This review provides information of great relevance to daily clinical practice. In spite of the number of studies performed to measure adherence, not all of the questionnaires used in the studies reach the requirements to be considered valid and reliable tools. Six questionnaires were identified to measure adherence to pharmacological antihypertensive treatment that had at least one validity test and one test of reliability. While some of these questionnaires had evidence to demonstrate acceptable validity, they failed in reliability and vice versa. Therefore although none of the questionnaires can be considered as a gold standard, this revision will assist health professionals in the selection of the most appropriate tool, depending on their circumstances. In the future, the design and validation of a questionnaire to measure adherence to antihypertensive treatments reaching the following requirements is fundamental: a) to be succinct enough to avoid patient and/or administrator fatigue (acceptability) but comprehensive enough to explain variance; b) include explicit content validity, have construct validity in accordance with a logical and reasoned theoretical structure that justifies items and subscales, and/or demonstrate criterion validity and; c) be a tool that provides reproducible results (reliable).

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


