Comparison of two self-rating instruments for medication adherence assessment in hypertension revealed insufficient psychometric properties

Janka Koschack*, Gabriella Marx, Jörg Schnakenberg, Michael M. Kochen, Wolfgang Himmel

Department of General Practice, University of Göttingen, Humboldtallee 38, 37073 Göttingen, Germany

Accepted 2 June 2009

Abstract

**Objective:** In cases of insufficiently controlled blood pressure, it is important for practitioners to distinguish between “nonadherence” and “nonresponse” to antihypertensive drug treatment. A reliable and valid adherence measurement based on the patient’s self-report may be helpful in daily practice.

**Study Design and Setting:** In a primary care sample with 353 hypertensive patients, we applied two self-rating instruments to assess medication adherence (the “Hill-Bone Compliance to High Blood Pressure Therapy Scale” and Morisky’s “Self-Reported Measure of Medication Adherence”) and compared their psychometric properties.

**Results:** Both scales showed low acceptability and insufficiency to moderate internal consistency (Cronbach’s $\alpha = 0.25$ and 0.73, respectively). Their convergent validity as indexed by kappa $= 0.39$ could be judged as “fair” at best. Testing the power to predict blood pressure $>140/90$ mm Hg, both scales showed an accuracy of 57% and 62%, respectively. The positive likelihood, that is, the increase in likelihood of high blood pressure in cases of nonadherence was 1.00 and 1.32, respectively.

**Conclusion:** The use of both scales cannot be recommended. They showed considerable floor effects, and their ability to identify medication adherence was inconsistent for nearly every third patient. The power of both scales to predict uncontrolled blood pressure was essentially a chance. The underlying conceptual framework of medication adherence therefore needs to be rethought. © 2010 Elsevier Inc. All rights reserved.

**Keywords:** Medication adherence; Hypertension; Psychometrics; Reproducibility of results; Primary health care; Questionnaires

1. Introduction

If evidence-based and effective therapies do not show the desired effect, a lack of patient medication adherence is often considered the main reason for this failure. In the case of hypertensive patients, a considerable proportion thereof has insufficient blood pressure control [1], and many authors suggest that a frequent reason is low medication adherence [2].

In clinical research, medical event monitoring systems that electronically record every opening of a pill box and similar methods are regarded as the current “gold standard” for adherence assessment [3,4]. However, their use in clinical practice is limited by at least two reasons: (1) high cost and (2) an atmosphere of control that might be counterproductive to the role of the patient as an active and autonomous agent in health-care decisions [5]. With respect to the patient perspective, the definition and degree of medication adherence should be a matter of “patient decision making” [6] or “asserting control over one’s disorder” [7]. For practitioners, who must distinguish between “nonadherence” and “nonresponse” to antihypertensive treatment in daily practice [8], a short and valid adherence measurement based on the patient’s self-report may be helpful.

The two scales frequently used in medication adherence research for hypertension are the “Hill-Bone Compliance to High Blood Pressure Therapy Scale” (Hill-Bone Scale) [9] and the “Self-Reported Measure of Medication Adherence” (Morisky-Green Scale) [10]. These instruments seem to be appropriate not only in treatment studies but also to inform practitioners and facilitate a frank discussion with the patient about his or her medication adherence. Both scales are short and easy to answer. The validity and reliability of these instruments have been investigated in some earlier studies [10,11] but not with homogenous results. Moreover, to our knowledge both adherence scales have not been compared in the same study population.
2. Methods

2.1. Study design and data collection procedures

This study was part of the MedViP project (“Medizinische Versorgung in der Praxis”—Medical Care in General Practice). The study protocol was approved by the University of Göttingen Research Ethics Committee. The design and recruitment have been described in detail elsewhere [12]. In brief, general practitioners were invited to provide routinely collected electronic medical data. Electronic patient records were extracted via a standardized interface.

We identified pseudonymized patient codes with a documented diagnosis of hypertension made on the electronic patient record. With the help of the practice assistants, a study nurse checked every hypertensive patient as a possible study participant. The exclusion criteria were unconfirmed hypertension diagnosis, emergency visits, or practice visits made during times when the practitioner has been temporarily replaced by a locum, mental, or terminal disease, and difficulties with verbal communication. The patients who agreed to participate in this study received a questionnaire covering sociodemographic and clinical information, and the Hill-Bone Scale. To limit the organizational impact, we decided to call only every second patient from the list. The patients were telephone interviewed to assess the Morisky-Green Scale within 2 weeks after the questionnaire. Fig. 1 shows the recruitment flow.

2.2. Sociodemographic data and clinical characteristics

The study questionnaire generated sociodemographic data (gender, age, marital status, education, and actual occupational status) and clinical characteristics. The patients were asked for the dates of the hypertension diagnosis and the initiation of drug treatment. If the patient remembered, the results of the most recent blood pressure measurement were recorded. We also asked the patients whether they regularly self-monitored their blood pressure. The study participants were requested to give detailed information about their drug intake on a separate sheet.

2.3. Medication adherence

The Hill-Bone Scale [9] assesses patient behavior for three behavioral domains of hypertension treatment and comprises 14 questions that are summed up to subscales: “reduced sodium intake” (three items), “appointment keeping” (two items), and “medication taking” (nine items). Each item could be answered on a four-point scale, resulting in a score ranging from 9 (perfect adherence) to 36 points. The authors reported sufficient psychometric properties (internal consistency, constructive and prospective validity). In a subsequent study, Krousel-Wood et al. [11] found sufficient internal consistency and factorial construct validity only for the medication-taking subscale. They recommended assessing adherence using only this subscale (“Hill-Bone Scale Short Form [SF]”) and dichotomized the responses in “perfect adherence” (9 points) and “imperfect adherence” (>9 points).

In our study, we used the original scale with all 14 items, but analyzed the adherence as suggested by the authors.

The Morisky-Green Scale [10] comprises four questions with a yes/no answer format. The resulting score ranges from 0 to 4 points, and the authors suggested a definition of high (0 points), medium (1–2 points), and low adherence (3–4 points). They reported sufficient internal consistency and criterion (predictive) validity. Medication adherence was assessed at baseline and blood pressure control was followed-up in a cohort of 290 patients over a 42-month time period. The authors showed a higher rate of controlled blood pressure in patients with high adherence. In our study, we used a dichotomous definition (perfect adherence = 0 points; imperfect adherence >0 points) because of the distribution of the scores with less than 2% of study participants having 3 or 4 points, that is, low adherence.

As the German translation did not exist for either the Hill-Bone Scale or the Morisky-Green Scale, we generated German language versions of these instruments using a modified forward/backward translation procedure, according to the protocol of the International Quality of Life Assessment project [13].
2.4. Statistical analyses

All analyses were performed using SPSS 16.0 (Statistical Package for the Social Sciences; SPSS Inc., Chicago, IL). The significance of group differences was tested by $t$-tests for metric variables and by chi-square tests for frequencies, as appropriate, with $P < 0.05$ (two-tailed) considered significant. Psychometric properties of the Hill-Bone Scale and the Morisky-Green Scale were evaluated in terms of acceptability, reliability, and validity.

2.4.1. Acceptability

We evaluated acceptability by examining response distributions, focusing on the maximum endorsement frequencies, that is, highest proportion of the respondents who endorsed a single category for an item (should be $<80\%$).

2.4.2. Reliability: internal consistency

Internal consistency refers to the extent to which items of the scale measure the same construct (i.e., homogeneity of the scale) and was assessed in our study by Cronbach’s $\alpha$ (should be $>0.70$) and the total item correlations (should be $>0.20$).

2.4.3. Construct validity

To verify construct validity, the factorial design of the Hill-Bone Scale was analyzed in a stepwise procedure. First, we tested the factorability of the intercorrelation matrix of the 14 items according to the Kaiser–Meyer–Olkin (KMO) coefficient (should be $>0.70$). In the second step, we conducted a principal component analysis (PCA) to derive an initial solution. Third, we determined the number of factors to be extracted according to three different criteria: (1) $eigenwert > 1.0$ [14] vs. (2) Cattell’s scree plot [15] vs. (3) the number of factors is identical with the proposed number of subscales (=3). In the last step, we compared the unrotated vs. the rotated factor solutions. The rationale of rotating factors is to obtain a simple factor structure that is more easily interpreted and compared. We chose the varimax rotation as the most popular method of orthogonal rotation. Each factor will tend to have either large or small loadings of any particular variable [16]. A factor loading, that is, correlation of an item with a factor, had to be $>0.40$ to be considered relevant. Because the Morisky-Green Scale contains just four items, which are summed up to one global score, it was not factorially analyzed.
2.4.4. Convergent validity

We tested the convergent validity of the Hill-Bone Scale SF and the Morisky-Green Scale using Cohen’s kappa coefficient as an assessment of the agreement “between two imperfect measurements of a variable subject to misclassification” [17]. The kappa statistic takes into account the agreement by chance. However, kappa as a measure of significance and its interpretation is controversial.

Although Landis and Koch have proposed standards for the strength of agreement, for example a “moderate” agreement if kappa ranges between 0.41 and 0.60 [18], other authors consider these benchmarks “inevitably arbitrary,” because kappa is affected by the number of categories in the measurement scale, the prevalence of the attribute, and the extent to which the raters (tests) disagree on the proportion of positive and negative cases [19,20]. In our study, we report not only the kappa coefficient, but also the raw data in a cross tab as suggested by Brennan and Silman [19]. Moreover, it would not have been reasonable to test the empirical kappa value in our study against kappa = 0, because a priori their agreement has to be better than expected by chance. The 95% confidence interval that indicates the range of plausible values for the “true” value of kappa should be more appropriate.

2.4.5. Criterion validity

To assess the criterion validity, we calculated measures of test accuracy for both medication adherence scales to determine their power to predict sufficiently controlled vs. uncontrolled blood pressure.

3. Results

The statistical analyses of the psychometric properties of the two medication adherence scales were based on 353 patients (from 23 general practices) who received drug treatment for their arterial hypertension. Table 1 provides a detailed description of the study sample.

3.1. Acceptability

The response frequencies for the nine items of the Hill-Bone Scale SF are listed in Table 2. The distribution of the four response options was biased toward the answer “none of the time.” Only the item “How often do you forget to take your HBP [high blood pressure] medicine?” had a tolerable maximum endorsement frequency of 71%. The observed range of the total score was 9–18, compared with a theoretical maximum score of 36, with just one participant having 18 points, but 215 participants having 9 points. The response frequencies of the four items of the Morisky-Green Scale are also listed in Table 2. The distribution of the dichotomous response option was biased to the answer “no.” None of the four items showed a tolerable maximum endorsement frequency of <80%. The four items could be summed up to the total score that leads to possible values ranging from 0 to 4. The observed range, however, was 0–3.

3.2. Internal consistency

For the Hill-Bone Scale SF, Cronbach’s α was 0.73. The total item correlations ranged from 0.00 to 0.65 (mean, 0.33) with one item failing the criterion of >0.20. For the Morisky-Green Scale, Cronbach’s α was 0.25. The total item correlations ranged from −0.01 to 0.19 (mean, 0.13) with all four items failing the criterion of >0.20.

3.3. Construct validity

The KMO coefficient for the Hill-Bone Scale was 0.73, confirming a sufficient degree of common variance and the factorability of the intercorrelation matrix of the 14 items. The initial solution after PCA revealed six components with eigen value >1.0. However, the scree plot begins to level off after three components, with a decrease of the eigen values from 1.44 to 1.17, consistent with the number of subscales. Therefore, we preferred the three-factor solution with eigen value (% variance explained) of 3.35 (24%), 1.56 (11%), and 1.44 (10%). Seven of the nine items of the medication-taking subscale and only one item of the appointment-keeping subscale loaded with >0.40 on the first factor. The remaining items loaded on the second and third factors without any meaningful pattern (Table 3; unrotated factor matrix). The varimax rotation
Table 2

Frequencies of responses for each item of the Hill–Bone Scale Sort Form (SF) and the Morisky-Green Scale

<table>
<thead>
<tr>
<th>Medication adherence scale</th>
<th>Response categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hill-Bone Scale SF (n = 353)</td>
<td>None of the time (1 point)</td>
</tr>
<tr>
<td>“How often do you forget to take your HBP medicine?” (%)</td>
<td>71</td>
</tr>
<tr>
<td>“How often do you decide not to take your HBP medicine?” (%)</td>
<td>91</td>
</tr>
<tr>
<td>“How often do you forget to get prescription refilled?” (%)</td>
<td>95</td>
</tr>
<tr>
<td>“How often do you run out of HBP pills?” (%)</td>
<td>94</td>
</tr>
<tr>
<td>“How often do you do it?” (%)</td>
<td>89</td>
</tr>
<tr>
<td>“How often do you miss taking your HBP pills when you feel better?” (%)</td>
<td>92</td>
</tr>
<tr>
<td>“How often do you take someone else’s HBP pills?” (%)</td>
<td>100</td>
</tr>
<tr>
<td>“How often do you miss taking your HBP pills when you are careless?” (%)</td>
<td>81</td>
</tr>
</tbody>
</table>

Morisky-Green Scale (n = 353)

| “Do you ever forget to take your medicine?” (%) | No (0 points) | Yes (1 point) |
| “Are you careless at times about taking your medicine?” (%) | 80 | 20 |
| “When you feel better do you sometimes stop taking your medicine?” (%) | 97 | 3 |
| “Sometimes you fell worse, when you take the medicine, do you stop taking it?” (%) | 96 | 4 |

Abbreviations: HBP, high blood pressure.

did not clarify the factor pattern or even deteriorated in case of the first factor (Table 3; rotated factor matrix).

3.4. Convergent validity

Table 4 shows the distribution of concordant positive or negative cases and of discordant cases. Of the 353 patients, 71% were classified concordantly, that is, 81 concordant negative cases and 169 concordant positive cases. Twenty-nine percent were classified discordantly, that is, 46 cases with nonadherence according to the Hill–Bone Scale SF but adherence according to the Morisky-Green Scale, and 57 cases vice versa. The observed kappa was 0.38.

3.5. Criterion validity

The participants were asked for their most recently measured systolic and diastolic blood pressure levels. We split the sample into patients who reported blood pressure rates <140/90 mm Hg (“sufficiently controlled”) vs. ≥140/90 mm Hg (insufficiently controlled) and tested the accuracy of the Hill–Bone Scale SF and the Morisky-Green Scale for predicting an insufficiently controlled blood pressure. Table 5 shows the true positives cases, that is, nonadherent patients with insufficiently controlled blood pressure, and the true negative cases, that is, adherent patients with sufficiently controlled blood pressure. The false positive cases, that is, nonadherent patients with sufficiently controlled blood pressure, and vice versa (=false negative cases) are also shown. The sensitivity and specificity of the Hill–Bone Scale SF (Morisky-Green Scale) were 37% and 63% (43% and 68%), respectively, with an accuracy barely above chance. Both scales showed a positive and negative likelihood with prevalence independent measures of test accuracy about 1.

Table 3

The three-factor solution of the Hill–Bone Scale

<table>
<thead>
<tr>
<th>Items of Hill–Bone Scale*</th>
<th>Unrotated matrix</th>
<th>Rotated matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factor 1</td>
<td>Factor 2</td>
</tr>
<tr>
<td>“How often do you forget to take your HBP medicine?”</td>
<td>0.56</td>
<td>−0.31</td>
</tr>
<tr>
<td>“How often do you decide not to take your HBP medicine?”</td>
<td>0.42</td>
<td>−0.37</td>
</tr>
<tr>
<td>“How often do you forget to get prescription refilled?”</td>
<td>0.59</td>
<td>0.50</td>
</tr>
<tr>
<td>“How often do you run out of HBP pills?”</td>
<td>0.61</td>
<td>0.46</td>
</tr>
<tr>
<td>“How often do you do it?”</td>
<td>0.32</td>
<td>0.05</td>
</tr>
<tr>
<td>“How often do you miss taking your HBP pills when you feel better?”</td>
<td>0.77</td>
<td>−0.35</td>
</tr>
<tr>
<td>“How often do you miss taking your HBP pills when you feel sick?”</td>
<td>0.68</td>
<td>−0.30</td>
</tr>
<tr>
<td>“How often do you take someone else’s HBP pills?”</td>
<td>−0.02</td>
<td>−0.02</td>
</tr>
<tr>
<td>“How often do you miss taking your HBP pills when you are careless?”</td>
<td>0.73</td>
<td>−0.32</td>
</tr>
<tr>
<td>“How do you eat salty food?”</td>
<td>0.32</td>
<td>0.44</td>
</tr>
<tr>
<td>“How often do you take your HBP medicine before you go to the doctor?”</td>
<td>0.29</td>
<td>0.39</td>
</tr>
<tr>
<td>“How often do you decide not to take your HBP medicine?”</td>
<td>0.21</td>
<td>−0.09</td>
</tr>
<tr>
<td>“How often do you make the next appointment before you leave the doctor’s office?”</td>
<td>0.15</td>
<td>0.01</td>
</tr>
<tr>
<td>“How often do you miss scheduled appointments?”</td>
<td>0.44</td>
<td>0.45</td>
</tr>
</tbody>
</table>

* Correlations between variables and factors >0.40 are bold; variables are sorted by subscales “medication taking,” “reduced sodium intake,” and “appointment keeping.”
4.1. Acceptability

The scores of both the Hill-Bone Scale SF and the Morisky-Green Scale showed considerable floor effects, with a strong tendency to opt for answers that represent “adherence.” This was particularly extreme for the Hill-Bone Scale, with all the respondents having total scores in the upper third of the possible distribution. It is unlikely that this skewed distribution reflects characteristics of our study population. Instead, it is more likely that the items themselves caused this bias by simplifying complex behavioral strategies and limiting answer variations. Cleopas et al. [21] showed that the acceptability and psychometric performance of a self-report instrument could be improved by using a five-point frequency response scale instead of a dichotomous response format. For example, the item “Do you ever forget to take your medicine?” of the Morisky-Green Scale with a dichotomous yes/no answer may be too rigorous. An interpretation of a “yes” as non-adherence would be going much too far, for example, a patient who has been treated for 10 years and forgot to take the tablet a single time only would be labeled as non-adherent. Vice versa, the interpretation of “no” as perfect adherence does not consider the possibility of a socially desired answer [22]. This may especially matter if adherence is assessed in medical contexts, for example, in a face-to-face interview with the doctor in person.

4.2. Internal consistency and convergent validity

Our factor analysis of the Hill-Bone Scale confirmed previous findings that the only meaningful factor that could be interpreted was the medication taking subscale [9,11]. Thus, both the Morisky-Green Scale and the Hill-Bone Scale could be reduced to one scale, implying that behavior can be assessed by an unidimensional scale. But the minimal levels of internal consistency of both instruments indicate that “adherence” is a more complex construct. The disappointing convergent validity of the two scales, which can be judged as “fair” at best, is another indication that a unidimensional construct, together with a simplistic operationalization by both scales, fails to consider “adherence” appropriately and therefore fails to measure it.

4.3. Criterion validity

The power of both instruments to predict controlled blood pressure was low. There was no change or just minimal increase in the likelihood of high blood pressure in case of nonadherence, as determined by both scales, and no change or just minimal decrease in the likelihood of high blood pressure in case of adherence, respectively. Before we consider this outcome as a failure of the two instruments, we should first discuss two alternative explanations, resulting from possible limitations of our study: (1) the representativeness of the study sample and (2) the reliability of self-reported blood pressure measurements.

1. Reliability of self-reported blood pressure measurements. In our study, the reported blood pressure level was based on regular self-monitoring in 65% of the cases (see Table 1). The validity of self-monitored blood pressure measures may be questioned. Home blood pressure measurement has become an accepted alternative to office-based measurement. It has been shown to be less influenced by observer bias and the “white coat effect” [23]. Studies have also revealed that the power of home blood pressure measurement to predict target-organ damage is comparable to that of the “gold standard,” that is, ambulatory monitoring [24]. Self-monitoring is also discussed as an act of “patient empowerment” that

Table 4
Comparison of the two scales

<table>
<thead>
<tr>
<th>Compliant acc. to</th>
<th>Hill-Bone Scale Short Form</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morisky-Green Scale</td>
<td>81</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td></td>
<td>46</td>
<td>169</td>
<td></td>
</tr>
</tbody>
</table>

Kappa = 0.38 (95% CI 0.28-0.49).

Table 5
Test accuracy of the two scales for predicting blood pressure ≥140/90 mm Hg by noncompliance (n = 325)

<table>
<thead>
<tr>
<th></th>
<th>True positive</th>
<th>True negative</th>
<th>False positive</th>
<th>False negative</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>LR+</th>
<th>LR−</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hill-Bone Scale SF</td>
<td>8</td>
<td>49</td>
<td>29</td>
<td>14</td>
<td>57</td>
<td>37</td>
<td>63</td>
<td>0.23</td>
<td>0.78</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Morisky-Green Scale</td>
<td>10</td>
<td>53</td>
<td>25</td>
<td>13</td>
<td>62</td>
<td>43</td>
<td>68</td>
<td>0.28</td>
<td>0.80</td>
<td>1.32</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Abbreviations: PPV, positive predictive value; NPV, negative predictive value; LR+, likelihood ratio for positive test result; LR−, likelihood ratio for negative test result.

All data given in % except for PPV, NPV, LR+, and LR−.
is correlated with better outcomes [25,26]. Although self-measurement of blood pressure seems to be valid and to have beneficial effects, the reliability of self-reported blood pressure measurements may be a separate matter for concern. Johnson et al. [27] showed that self-reported blood pressure readings were identical to electronic stored measurements in more than 80% of the cases; erroneous reporting occurred more frequently in patients with elevated blood pressure than in those with controlled blood pressure. Thus, we cannot exclude that the self-reported blood pressure measurements in our study were inappropriate in some patients and in particular too optimistic.

2. Representativeness of the study sample. The response rate in our study was only 28% and the final sample consisted of 353 patients (see Fig. 1), so that only every 10th eligible patient in the initial population was included and provided data. However, the detected rates of nonadherence in our study, 31% (Hill-Bone Scale SF) and 36% (Morisky-Green Scale), respectively (see Table 1), were comparable to the nonadherence rate of 30% reported by Krousel-Wood et al. [11] and a systematic review [28]. Morisky et al. [10] reported a somewhat higher rate of nonadherence (57%). Although the rates of adherence are similar, our study sample may have been biased, such that the rate of patients being nonadherent was different in the initial population than in the study sample and nonrespondents might be even more likely to be nonadherent. Despite this possible selection bias, our study still detected the insufficient predictive power of the two scales; namely that out of four patients with low adherence as determined by both scales only one has a reported blood pressure level of $\geq 140/90$ mm Hg.

4.4. Conclusion

Although some of the limitations of the psychometric properties of the scales may have been increased by the possible bias of our study, the simplistic assessment of medication intake behavior seemed to be the main reason why both instruments did not meet the standards of acceptability, reliability, and validity. Both the Hill-Bone Scale and the Morisky-Green Scale clearly failed to help physicians differentiate between “nonadherence” and “nonresponse” to antihypertensive drug treatment. Other methods such as electronic monitoring devices proved to be effective if drug intake behavior alone is studied in clinical trials [29]. However, in primary care and other clinical settings the patient’s strategies to cope with a chronic disease and the “chronic” drug taking in his or her everyday life seem to be an even more important aspect of medication adherence. To assess and discuss these strategies, better self-ratings and questionnaires—rather than monitoring devices—are needed.

Overall, we concur with Lee Sechrest’s call “to think more about, do more about, and write more about the validity of the data we produce and less about the validity of specific instruments” [30]. Medication adherence is however influenced at many levels, for example, by behavior (e.g., forgetting to take the pill), emotion (e.g., being afraid of long-term effects of drug intake), and cognition (e.g., deciding not to take the diuretic so as to enjoy Richard Wagner’s 5-hour opera “Parsifal” uninterrupted). Therefore maybe instead of advanced optimization of existing scales, we suggest that the conceptual framework of medication adherence should be rethought.

Acknowledgments

We thank all patients who participated in this study for their cooperation. The assistance of Hannelore Schneider-Ruth in collecting the data is gratefully acknowledged.

References