

**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
Göttingen, Germany

Corresponding Author

Dr Janka Koschack

Department of General Practice

University of Göttingen

Humboldtallee 38

37073 Göttingen, Germany

Phone: +49 (0)551 39 14221

Fax: +49 (0)551 39 14222

jkoscha@gwdg.de

Abstract

Objective

In cases of insufficiently controlled blood pressure, it is important for practitioners to distinguish between 'non-adherence' and 'non-response' 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 Settings

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 insufficient 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 $>140/90$ mmHg, both scales showed an accuracy of 57% and 62%, respectively. The positive likelihood, i.e. the increase in likelihood of high blood pressure in cases of non-adherence, 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 chance. The underlying conceptual framework of medication adherence therefore needs to be rethought.

Keywords: medication adherence; hypertension; psychometrics; reproducibility of results; primary health care; questionnaires

Running Title: Comparison of two medication adherence scales for hypertension

Word Count: 3267

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 have 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 'non-adherence' and 'non-response' to antihypertensive treatment in daily practice [8], a short and valid adherence measurement based on the patient's self-report may be helpful.

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.

The aim of our study was to test the acceptability, reliability and validity of the *Hill-Bone-Scale* and the *Morisky-Green-Scale*. In particular, we assessed the convergent validity of both instruments and the criterion validity, i.e. the strength of association between the scales and the quality of blood pressure control as reported by the patients themselves.

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. 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 standardised interface.

We identified pseudonymised 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. 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. Patients who agreed to participate in this study received a questionnaire covering sociodemographic and clinical information and the *Hill-Bone-Scale*. To limit the organisational impact, we decided to call only every second patient from the list. Patients were telephone interviewed to assess the *Morisky-Green-Scale* within two weeks after the questionnaire. Figure 1 shows the recruitment flow.

Insert figure 1 here

2.2. Sociodemographic data and clinical characteristics

The study questionnaire generated sociodemographic data (gender, age, marital status, education, and actual occupational status) as well as clinical characteristics.

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. 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 behaviour for three behavioural 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 4-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 and colleagues [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 dichotomised 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 analysed 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 to 2 points), and low adherence (3 to 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) due to the distribution of the scores with less than 2% of study participants having 3 or 4 points i.e. low adherence.

As a 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 (IQOLA) project [13].

2.4. Statistical analyses

All analyses were performed using SPSS 16.0 (Statistical Package for the Social Sciences) (SPSS Inc., Chicago, IL 60606). The significance of group differences was tested by *t*-tests for metric variables and by *chi*²-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.

Acceptability

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

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 α (should be >0.70) and the total-item correlations (should be >0.20).

Construct validity

To verify construct validity, the factorial design of the *Hill-Bone-Scale* was analysed 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 a 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 solution. 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, i.e. 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, that are summed up to one global score, it was not factorially analysed.

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.

While 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, 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 [18,19]. 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 [18]. Moreover, it would not have been reasonable to test the empirical kappa value in our study against $\text{kappa}=0$, 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.

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.

Insert table 1 here

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 towards the answer “none of the time”. Only the item “How often do you forget to take your HBP medicine?” had a tolerable maximum endorsement frequency of 71%. The observed range of the total score was 9 to 18, compared to a theoretical maximum score of 36, with just one participant having 18 points, but 215 participants having nine 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 to 3.

Insert table 2 here

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 eigenvalue >1.0 . However, the scree plot begins to level off after three components, with a decrease of the eigenvalues from 1.44 to 1.17, consistent with the number of subscales. Therefore, we preferred the three-factor solution with eigenvalue (% 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 factor without any meaningful pattern (Table 3; unrotated factor matrix). The varimax rotation did not clarify the factor pattern or even deteriorated in case of the first factor (Table 3; rotated factor matrix).

Insert table 3 here

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, i.e. 81 concordant negative cases and 169 concordant positive cases. Twenty nine percent were classified discordantly, i.e. 46 cases with non-adherence 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.

Insert table 4 here

3.5. Criterion validity

Participants were asked for their most recently measured systolic and diastolic blood pressure level. We split the sample into patients reported blood pressure rates $<140 / 90$ mm Hg ('sufficiently controlled') vs. $\geq 140 / 90$ 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, i.e. non-adherent patients with insufficiently controlled blood pressure, and the true negative cases, i.e. adherent patients with sufficiently controlled blood pressure. The false positive cases, i.e. non-adherent 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.

Insert table 5 here

4. Discussion

Based on a sample of 353 patients with drug treated hypertension, we examined the psychometric properties of two self-rating instruments for medication adherence, the '*Hill-Bone Compliance to High Blood Pressure Therapy Scale*' [9] and the '*Self Reported Measure of Medication Adherence*' [10]. Both instruments showed an insufficient psychometric quality in different aspects of acceptability, reliability and validity.

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 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 behavioural strategies and limiting answer variations. Cleopas and colleagues [20] showed that the acceptability and psychometric performance of a self-report instrument could be improved by using a 5-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, e.g. a patient who has been treated for ten years and forgot to take the tablet a single time only would be labelled as non-adherent. Vice versa, the interpretation of 'no' as perfect adherence does not consider the possibility of a socially desired answer [21]. 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 behaviour can be assessed by an unidimensional scale. But the minimal levels of internal consistency of both instruments indicates 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 operationalisation 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 non-adherence, 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' [22]. 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', i.e. ambulatory monitoring [23]. Self-monitoring is also discussed as an act of 'patient empowerment' that is correlated with better outcomes [24,25]. 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 and colleagues [26] 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 figure 1), so that only every tenth eligible patient in the initial population was included and provided data. However, the detected rates of non-adherence in our study, 31% (*Hill-Bone-Scale SF*) and 36% (*Morisky-Green-Scale*), respectively (see table 1), were comparable to the non-adherence rate of 30% reported by Krousel-Wood and colleagues [11] and a systematic review [27]. Morisky and colleagues [10] reported a somewhat higher rate of non-adherence (57%). Although the rates of adherence are similar, our study sample may have been biased, such that the rate of patients being non-adherent was different in the initial population than in the study sample and non-respondents might be even more likely to be non-adherent. 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 $\geq 140/90$ 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 behaviour 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 ‘non-adherence’ and ‘non-response’ to antihypertensive drug treatment. Other methods like electronic monitoring devices proved to be effective if drug intake behaviour alone is studied in clinical trials [28]. 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” [29]. Medication adherence is however influenced at many levels, for example, by behaviour (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 five hour opera ‘Parsifal’ uninterrupted). Therefore maybe instead of advanced optimisation of existing scales, we suggest that the conceptual framework of medication adherence should be rethought.

Acknowledgements

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

- [1] Wolf-Maier K, Cooper RS, Kramer H, Banegas JR, Giampaoli S, Joffres MR et al. Hypertension treatment and control in five European countries, Canada, and the United States. *Hypertension* 2004;43:10-7.
- [2] Burnier M. Medication adherence and persistence as the cornerstone of effective antihypertensive therapy. *Am J Hypertens* 2006;19:1190-6.
- [3] Farmer KC. Methods for measuring and monitoring medication regimen adherence in clinical trials and clinical practice. *Clin Ther* 1999;21:1074-90.
- [4] Urquhart J. The electronic medication event monitor. Lessons for pharmacotherapy. *Clin Pharmacokinet* 1997;32:345-56.
- [5] Sullivan M. The new subjective medicine: taking the patient's point of view on health care and health. *Soc Sci Med* 2003;56:1595-604.
- [6] Donovan JL. Patient decision making. The missing ingredient in compliance research. *Int J Technol Assess Health Care* 1995;11:443-55.
- [7] Conrad P. The meaning of medications: another look at compliance. *Soc Sci Med* 1985;20:29-37.
- [8] Zeller A, Schroeder K, Peters TJ. An adherence self-reported questionnaire facilitated the differentiation between nonadherence and nonresponse to antihypertensive treatment. *J Clin Epidemiol* 2008;61:282-8.
- [9] Kim MT, Hill MN, Bone LR, Levine DM. Development and testing of the Hill-Bone Compliance to High Blood Pressure Therapy Scale. *Prog Cardiovasc Nurs* 2000;15:90-6.

- [10] Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care* 1986;24:67-74.
- [11] Krousel-Wood M, Muntner P, Jannu A, Desalvo K, Re RN. Reliability of a medication adherence measure in an outpatient setting. *Am J Med Sci* 2005;330:128-33.
- [12] Wetzel D, Himmel W, Heidenreich R, Hummers-Pradier E, Kochen MM, Rogausch A et al. Participation in a quality of care study and consequences for generalizability of general practice research. *Fam Pract* 2005;22:458-64.
- [13] Bullinger M, Alonso J, Apolone G, Lepège A, Sullivan M, Wood-Dauphinee S et al. Translating health status questionnaires and evaluating their quality: the IQOLA project approach. *J Clin Epidemiol* 1998;51:913-23.
- [14] Guttman L. Some necessary conditions for common-factor analysis. *Psychometrika* 1954;19:149-61.
- [15] Cattell RB. The scree test for the number of factors. *Multivariate Behav Res* 1966;1:245-76.
- [16] Kaiser HF. The VARIMAX criterion for analytic rotation in factor analysis. *Psychometrika* 1958;23:187-200.
- [17] Thompson WD, Walter SD. A reappraisal of the kappa coefficient. *J Clin Epidemiol* 1988;41:949-58.
- [18] Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ* 1992;304:1491-4.

- [19] Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257-68.
- [20] Cleopas A, Kolly V, Perneger TV. Longer response scales improved the acceptability and performance of the Nottingham Health Profile. *J Clin Epidemiol* 2006;59:1183-90.
- [21] Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm* 2008;65:2276-84.
- [22] Verberk WJ, Kroon AA, Kessels AGH, de Leeuw PW. Home blood pressure measurement. A systematic review. *J Am Coll Cardiol* 2005;46:743-51.
- [23] Stergiou GS, Argyraki KK, Moyssaki I, Mastorantonakis SE, Achimastos AD, Karamanos VG et al. Home blood pressure is as reliable as ambulatory blood pressure in predicting target-organ damage in hypertension. *Am J Hypertens* 2007;20:616-21.
- [24] Figar S, Galarza C, Petrlik E, Hornstein L, Rodríguez Loria G, waisman G et al. Effect of education on blood pressure control in elderly persons: a randomized controlled trial. *Am J Hypertens* 2006;19:737-43.
- [25] Halme L, Vesalainen R, Kaaja M, Kantola I. Self-monitoring of blood pressure promotes achievement of blood pressure target in primary health care. *Am J Hypertens* 2005;18:1415-20.
- [26] Johnson KA, Partsch DJ, Rippole LL, McVey DM. Reliability of self-reported blood pressure measurements. *Arch Intern Med* 1999;159:2689-93.

- [27] DiMatteo MR. Variations in patients' adherence to medical recommendations. A quantitative review of 50 years of research. *Med Care* 2004;42:200-9.
- [28] Wetzels GEC, Nelemans PJ, Schouten JSAG, van Wijk BLG, Prins MH. All that glitters is not gold: a comparison of electronic monitoring versus filled prescriptions - an observational study. *BMC Health Serv Res* 2006;6:8.
- [29] Sechrest L. Validity of measures is no simple matter. *Health Serv Res* 2005;40:1584-604.

Legends

Figure 1 Flow chart of the study sample recruitment

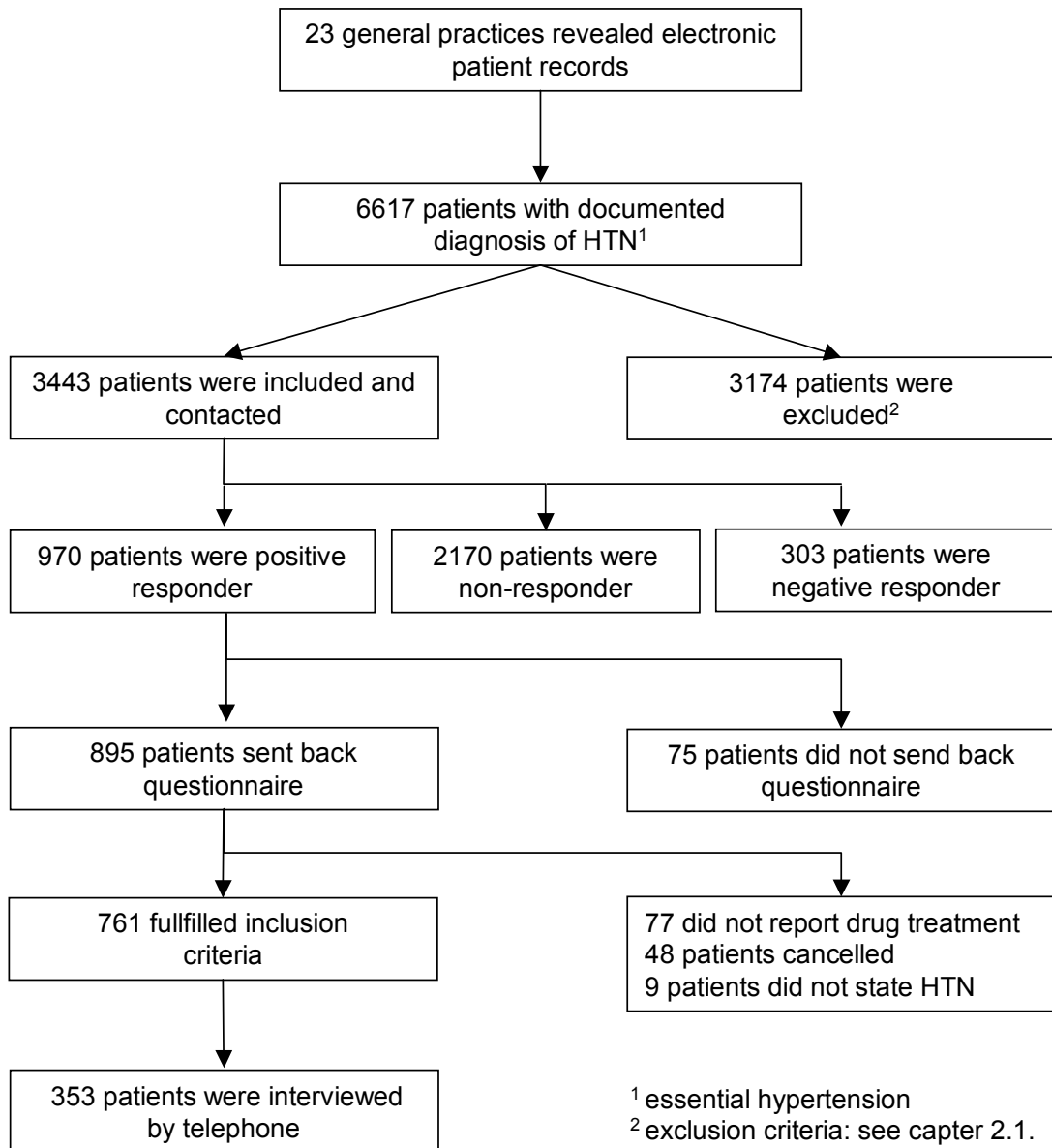


Table 1 Description of the study sample (n=353)

Sociodemographic data	
Age in years	64 (11)
Male, %	51
Higher degree, %	25
Employed, %	33
Living alone, %	22
Clinical characteristics	
Systolic BP, mm Hg	138 (15)
Diastolic BP, mm Hg	82 (11)
BP >140/90 mm Hg, %	23
Drug therapy >5 years, %	55
Regular self-monitoring, %	63
Medication adherence	
<i>Hill-Bone-Scale SF</i> , n of adherent, %	61
<i>Morisky-Green-Scale</i> , n of adherent, %	64
Prescribed cardiovascular drugs	
Beta blocker, %	53
ACE inhibitor/ Angiotension II blocker, %	77
Calcium channel blocker, %	24
Diuretic ¹ , %	45
Alpha blocker, %	7
Digitalis/ Nitrate , %	7
N of prescribed cardiovascular drugs > 2, %	40

Table 2 Frequencies of responses for each item of the *Hill-Bone-Scale SF* and the *Morisky-Green-Scale*

Medication adherence scale	Response categories			
<i>Hill-Bone-Scale SF</i> (n=353)	none of the time (1 pts.)	some of the time (2 pts.)	most of the time (3 pts.)	all of the time (4 pts.)
“How often do you forget to take your HBP medicine?”	71%	29%	0%	0%
“How often do you decide not to take your HBP medicine?”	91%	8%	0%	1%
“How often do you forget to get prescription refilled?”	95%	5%	0%	0%
“How often do you run out of HBP pills?”	94%	6%	0%	0%
“How often do you skip your HBP medicine before you go to the doctor?”	94%	6%	0%	0%
“	88%	11%	1%	0%
“	92%	8%	0%	0%
“How often do you take someone else’s HBP pills?”	100%	0%	0%	0%
“How often do you miss taking your HBP pills when you are careless?”	81%	9%	0%	0%
<i>Morisky-Green-Scale</i> (n=353)	no (0 pts.)	yes (1 pts.)		
“Do you ever forget to take your medicine?”	80%	20%		
“Are you careless at times about taking your medicine?”	81%	19%		
“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%		

Table 3 The three factor solution of the *Hill-Bone-Scale*

Items of <i>Hill-Bone-Scale</i> ¹	Unrotated matrix			Rotated matrix		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
“How often do you forget to take your HBP medicine?”	.56	-.31	.47	.68	.06	.41
“How often do you decide not to take your HBP medicine?”	.42	-.37	-.37	.52	-.04	-.43
“How often do you forget to get prescription refilled?”	.59	.50	-.17	.17	.75	-.17
“How often do you run out of HBP pills?”	.61	.46	-.20	.21	.73	-.20
“How often do you skip your HBP medicine before you go to the doctor?”	.32	.05	-.52	.17	.24	-.53
“	.77	-.35	.01	.82	.17	-.07
“	.68	-.30	-.21	.70	.16	-.28
“How often do you take someone else’s HBP pills?”	-.02	-.02	.10	.01	-.02	.10
“How often do you miss taking your HBP pills when you are careless?”	.73	-.32	.21	.80	.16	.14
“How often do you eat salty food?”	.32	.44	.42	.05	.53	.43
“How often do you shake salt on your food before you eat it?”	.29	.39	.29	.04	.48	.31
“How often do you eat fast food?”	.21	-.09	.11	.24	.05	.08
“How often do you make the next appointment before you leave the doctor’s office?”	.15	.01	.58	.18	.08	.57
“How often do you miss scheduled appointments?”	.44	.45	-.16	.08	.63	-.15

¹ Correlations between variables and factors > 0.40 are bold; variables are sorted by subscales ‘medication taking’, ‘reduced sodium intake’, ‘appointment keeping’.

Table 4 Comparison of the two scales

		Compliant acc. to <i>Morisky-Green-Scale</i>		
		no	yes	total
Compliant acc. to <i>Hill-Bone-Scale SF</i>	no	81	57	138
	yes	46	169	215
	total	127	226	353

Table 5 Test accuracy of the two scales for predicting blood pressure $\geq 140 / 90$ mm Hg by non compliance (n=325)

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

All data given in % except for PPV, NPV, LR+, and LR-.

PPV positive predictive value; NPV negative predictive value; LR+ likelihood ratio for positive test result; LR- likelihood ratio for negative test result.

What is new?**Key findings**

Convergent validity, criterion validity, and accuracy of two self rating instruments to measure medication adherence in hypertension have been proved to be insufficient for use in clinical practice and research.

What this adds to what was known

Beside poor psychometric properties caused by too simplistic items and an arbitrary scoring procedure, these deficiencies also mirror shortcoming in conceptualizing adherence.

What is the implication, what should change now

The construction of new measurements of medication adherence should take the patient perspective more seriously as an important condition for sufficient psychometric properties.