

Psychometric Properties of the Dutch Version of the Hospital-Level Consumer Assessment of Health Plans Survey[®] Instrument

Onyebuchi A. Arah, A. H. A. ten Asbroek, Diana M. J. Delnoij, Johan S. de Koning, Piet J. A. Stam, Aldien H. Poll, Barbara Vriens, Paul F. Schmidt, and Niek S. Klazinga

Objectives. To assess the reliability and validity of a translated version of the American Hospital-level Consumer Assessment of Health Plans Survey[®] (H-CAHPS) instrument for use in Dutch health care.

Data Sources/Study Setting. Primary survey data from adults aged 18 years or more who were recently discharged from two multispecialty city hospitals in the Netherlands.

Study Design. We used forward and backward translation procedures and a panel of experts to adapt the 66-item pilot H-CAHPS into a 70-item Dutch instrument. Descriptive statistics and standard psychometric methods were then used to test the reliability and validity of the new instrument.

Data Collection. From late November 2003 to early January 2004, the survey was administered by mail to 1,996 patients discharged within the previous 2 months.

Principal Findings. Analyses supported the reliability and validity of the following 7-factor H-CAHPS structure for use in Dutch hospitals: on doctor's communication, nurses' communication, discharge information, communication about medication, pain control, physical environment of hospital, and nursing services. The internal consistency reliability of the scales ranged from 0.60 to 0.88. Items related to "family receiving help when on visit," "hospital staff introducing self," and "admission delays" did not improve the psychometric properties of the new instrument.

Conclusions. These findings suggest that the H-CAHPS instrument is reliable and valid for use in the Dutch context. However, more research will be needed to support its equivalence to the United States version, and its use for between-hospital comparisons.

Key Words. CAHPS[®], hospitals, Dutch health care, patient experiences of care, cross-cultural translation

Like in many advanced countries, current health care reforms which emphasize accountability, transparency, choice, and performance improvement encourage Dutch patients to become critical consumers (Ministry of Health, Welfare and Sport 2002; Ministerie van Volksgezondheid, Welzijn en Sport 2004; ten Asbroek et al. 2004). Health care seekers need all the help they can get to identify appropriate and high quality health care providers, insurers, and institutions (Iezzoni 2002). Surveys which assess patients' health care experiences are increasingly being used by consumers and insurers to choose and contract high performers, as well as to hold them accountable. Several Dutch instruments have been previously developed to capture generic, group-specific (e.g., elderly), or disease-specific (e.g., HIV) health care experiences, satisfaction, opinions, and preferences (Sixma et al. 1998, 2000; Sixma, Spreeuwenberg, and van der Pasch 1998; Hendriks et al. 2000, 2002; Jansen, Hutten, and Spreeuwenberg 2002; Hekkink et al. 2003). The growing role of cross-national learning and comparisons of health systems performance (Arah et al. 2003) imply that there are opportunities for standardization of survey instruments and for international comparisons.

After an extensive literature review, the United States Consumer Assessment of Health Plans Survey (CAHPS[®]) instruments were selected for adaptation and use in Dutch pilot studies. The Dutch health care system is seeking a family of related and standardized instruments such as the CAHPS[®] products for multiple (choice, purchasing, and performance improvement) purposes. The CAHPS[®] products are well-established in the American context (Hays et al. 1999; Spranca et al. 2000; Goldstein and Fyock 2001; Morales et al. 2001; Farley et al. 2002; Hargraves, Hays, and Cleary 2003; CAHPS[®] Survey Users Network [SUN] 2004; Solomon et al. 2005). Funded by the Centers for Medicare and Medicaid Services in partnership with the Agency for Healthcare Research and Quality in the United States (U.S.), the Hospital CAHPS[®] (H-CHAPS) instrument was recently developed and field tested to

Address correspondence to Onyebuchi A. Arah, M.D., D.Sc., Department of Social Medicine, Academic Medical Center of the University of Amsterdam, PO Box 22700, 1100 DE, Amsterdam, The Netherlands. Dr. Arah is also with the Center for Prevention and Health Services Research, National Institute of Public Health and the Environment, Bilthoven, The Netherlands. A. H. A. ten Asbroek, M.Sc., Johan S. de Koning, M.P.H., Ph.D., and Niek S. Klazinga, M.D., Ph.D. are with the Department of Social Medicine, Academic Medical Center of the University of Amsterdam, Amsterdam, The Netherlands. Diana M. J. Delnoij, Ph.D. is with The Netherlands Institute for Health Services Research (Nivel), Utrecht, The Netherlands. Piet J. A. Stam, M.Sc., Aldien H. Poll, Barbara Vriens, M.Sc., and Paul F. Schmidt are with the Agis Health Insurance Company, Amersfoort, The Netherlands. Piet J. A. Stam is also with the Institute for Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands.

provide comparative information on hospitals and to aid informed choices by patients (CAHPS Survey Users Network 2004; Centers for Medicare and Medicaid Services 2004). During the pilot study in the U.S. in 2003, we received and adapted the pilot version of the H-CAHPS for use in Dutch hospitals. This study is part of concerted efforts, initially spearheaded by a large Dutch health insurer, to import a family of CAHPS[®] instruments, including ambulatory care (such as the health plan) surveys and facility surveys (such as the H-CAHPS), for measuring health care experiences from the perspective of Dutch patients and consumers, and to use the resulting information for their health care purchasing purposes.¹ The aim of this study is to report on the reliability and validity of the H-CAHPS in Dutch hospitals.

METHODS

H-CAHPS Instrument and the Dutch Translation

We used the 2003 U.S. pilot version of the H-CAHPS: a 66-item instrument which contained 33 core items on patient experiences, three global ratings (of nurses, doctor, and the hospital), one item on the likelihood to recommend hospital to friends and family, and several items on patient characteristics. Twenty-seven of the core items on patient experiences were evaluated on a 1-to-4 response scale, where 1 referred to “never,” 2 “sometimes,” 3 “usually,” and 4 “always.” The remaining six core items had a dichotomous (yes–no) response scale. The global ratings had a 0–10 response scale with only the endpoints labeled (e.g., 0 was “worst hospital possible” and 10 was “best hospital possible”). The one question on patient’s likelihood to recommend the hospital to friends and family had the following response scale: 1 “definitely no,” 2 “probably no,” 3 “probably yes,” and 4 “definitely yes.” The instrument asked respondents to think about their last stay at the specified and confirmed hospital. The development of the H-CAHPS through extensive systematic literature review, consumer focus groups, public response to Federal Register notice, stakeholder input, cognitive testing, and a 3-state pilot test in New York, Arizona, and Maryland is detailed elsewhere (The CAHPS[®] II Investigators and the Agency for Healthcare Research and Quality 2003).

Mindful of the difficulties and equivalence issues involved in transplanting survey instruments from one culture to another (Weidmer, Brown, and Garcia 1999; Streiner and Norman 2003), we had H-CAHPS translated into Dutch by two independent professional translators, and subsequently

backtranslated into English by two other independent translators who had never seen the original U.S. version. A panel of seven researchers experienced in patient surveys or survey instrument development was asked to choose a mix of the two versions that came closest to the original U.S. instrument and used clear, comprehensible language. A few adaptations were also made to fit the Dutch context as follows: the two U.S. questions on Hispanic descent and race were replaced by three items on birth place of respondent, mother, and father; and three additional items on getting verbal information on activity limitations (Q47), getting help when home (Q50), and taking new medication at home (Q53). The new Dutch instrument totaled 70 items which included 35 core items.

The Pilot Survey and Subjects

The study sites were two multispecialty city hospitals with 555 and 386 beds, and 15,761 and 12,606 admissions, respectively, in 2003. Assuming a possible 50 percent response rate, a random sample of 998 patients per hospital was drawn to reflect the hospital's patient population aged 18 years and above and discharged from admission (lasting at least one night) within the previous 2 months. The survey was administered by mail from late November 2003 to early January 2004. Nonrespondents were followed up with a postcard 1 week later, a second questionnaire 3 weeks later, and a reminder letter 5 weeks later. Each mailing pack included a stamped addressed envelope and a cover letter using each hospital's letterhead, endorsed by the Dutch Patients and Consumers Federation, and guaranteeing confidentiality.

Data Analysis

We conducted several analyses to assess response and psychometric properties of the instrument. The respondents and nonrespondents could only be compared on age and sex because of Dutch privacy laws. We assessed the percentage of valid and invalid response/skip patterns, and the percentage of responses within the minimum and maximum response categories for each item. Using principal component analysis (with oblique rotation) we explored the factor structure. Principal factor analysis would yield similar results with 30 or more variables exhibiting high communalities as was the case here (Stevens 1992). If an item loaded across multiple factors, it was assigned to the factor where it had the highest loading and/or subsequently maximized the internal consistency.

The internal consistency reliability of the scales was estimated using Cronbach's α (Cronbach 1951), where an α value of 0.70 or more was considered satisfactory. Item-total scale correlations were also calculated, correcting for item overlap, to check for homogeneity of the simple-summed scales created from items that loaded strongly on the factors (Streiner and Norman 2003). Each of these correlations was also checked to see if it was greater than the correlation of each item with scales other than its own. Analyses using imputed values derived from a multivariate normal model (as implemented in SAS PROC MI, *SAS* Version 9.1) to replace missing data yielded similar results to the ones from the pairwise deletion methods reported here.

Using Pearson product moment correlations (for the ratings) and Spearman rank-order correlations (for the items), we also evaluated associations between (a) items, (b) simple-summed scales from the extracted factors, (c) scales and the global ratings (of nurse, doctor, and hospital) and the likelihood to recommend hospital. The interscale correlations above were to give further insights into the interpretability of the constructed factors as separate scales (if correlations were less than 0.70, a value that follows a similar logic as the scale reliability threshold), thus supporting the multidimensionality of the questionnaire (Carey and Seibert 1993). The correlations between scales and the global ratings were also evaluated using adjusted regression models.

Finally, we evaluated the associations between respondent characteristics and the three global ratings, using multiple linear regression models specified for all respondents, and separately for respondents who were admitted for surgical, childbirth, or other medical reasons. We also looked at the effects of respondent characteristics on patient experience composites, in order to aid future analysis of between-hospital differences (not reported here). All analyses were conducted using the *Statistical Package for the Social Sciences* (SPSS Inc. 2004).

RESULTS

Response and Survey Subjects

The survey response rates were 63 and 57 percent for hospitals A and B, respectively (averaging 60 percent, $n = 1,194$). The respondents were not significantly older than the nonrespondents (mean age 53.2 versus 52.3 years, $p = .32$). However, there were significantly more females among respondents than among nonrespondents (64 versus 61 percent, $p = .03$). Table 1 summarizes the respondents' characteristics.

Table 1: Characteristics of Survey Respondents

<i>Characteristics (Valid %)</i>	<i>Respondents (n = 1,194)</i>
Age (years)	
Mean (SD)	53.2 (20.0)
18–64 (%)	64.6
65+ (%)	35.4
Gender (%)	
Male	35.4
Female	64.6
Education (%)	
No or less than secondary education	50.4
Secondary or higher education	49.6
Birth country (%)	
The Netherlands	73.0
Language spoken at home (%)	
Dutch	85.5
Self-reported general health status (%)	
Excellent	6.6
Very good	10.4
Good	33.4
Fair	39.2
Poor	10.4
Self-reported mental health status (%)	
Excellent	17.2
Very good	13.4
Good	38.2
Fair	26.0
Poor	5.2
Reason for hospital admission (%)	
Surgical	46.0
Childbirth	14.3
Other medical	39.7
Discharge destination (%)	
Own home	91.5
Someone else's home	4.2
Another health care facility	4.2

SD, standard deviation.

The percentage of cases within the minimum response category value ranged from 0.4% for item Q4 (on how often nurses were respectful) to 66 percent for the dichotomous response item Q49 (on getting written information on problems to watch out for after discharge) (Table 2). For those within the maximum response category, the percentage ranged from 25 percent for item Q41 on being told side-effects of new medicine to 76 percent for item Q53 on getting verbal information on taking medicine at home. An analysis of

Table 2: Item and Scale Characteristics, Internal Consistency Reliability, and Item-Total Correlations

Dutch Item Number	Scale and Items	% Responses in Minimum/Maximum Response Category		Loading on Primary Factor/Scale	Internal Consistency Reliability: Cronbach's α	Corrected Item-Total Correlations
		per Item	per Item			
	<i>Doctor's care/communication*</i>				0.88	
Q11	Doctor respected	1.0/64.5		0.89		0.73
Q12	Doctor listened carefully	3.2/54.5		0.92		0.79
Q13	Doctor explained things clearly	4.9/50.2		0.79		0.71
Q14	Doctor spent enough time	4.8/41.4		0.84		0.75
Q25	Involved in own treatment decisions	8.8/37.8		0.48		0.57
	<i>Nurses' care/communication*</i>				0.84	
Q4	Nurses respected	0.4/68.5		0.90		0.68
Q5	Nurses listened carefully	1.9/53.8		0.85		0.74
Q6	Nurses explained things clearly	4.6/50.8		0.73		0.61
Q7	Nurses spent enough time	2.8/43.2		0.76		0.68
	<i>Discharge information†</i>				0.76	
Q46	Activities could do or not (written)	62.0/38.0		0.78		0.60
Q47	Activities could do or not (verbal)	36.8/63.2		0.62		0.54
Q48	Postdischarge help if needed	48.3/51.7		0.52		0.39
Q49	Problems to look out for (written)	66.0/34.0		0.79		0.53
Q50	Problems to look out for (verbal)	44.6/55.4		0.67		0.52
Q52	Taking medicine at home (written)	48.8/51.2		0.38		0.39
Q53	Taking medicine at home (verbal)	24.0/76.0		0.36		0.35
	<i>Communication about medication*</i>				0.81	
Q37	Told name of new medicine	19.3/39.3		0.79		0.65
Q38	Told what medicine was for	10.7/53.4		0.81		0.62
Q39	Asked if taking other medicines	13.7/53.7		0.62		0.56
Q40	Asked if allergic to any medicines	23.2/45.5		0.66		0.54
Q41	Told side effects of medicine	36.8/25.1		0.64		0.60

the response and skip patterns revealed that items were about 92 percent to more than 99 percent consistently completed.

Psychometric Properties: Factor Structure, Reliability, and Validity

Principal components analysis with oblique rotation yielded seven factors with eigenvalues greater than 1, and which explained 60 percent of the total variance among the items covering patient experiences. The scree test/plot and estimated communalities also lent support to the 7-factor structure. Only the primary factor loadings based on the pattern matrix are presented in Table 2. We dropped items Q27 (family receiving help when visiting), Q28 (hospital staff introducing themselves), and Q43 (admission delay) because they exhibited poor or unclear (difficult to interpret) factor loadings, and did not subsequently improve the internal consistency of the instrument. The remaining 32 items formed the following scales/composites: (1) doctor's communication, (2) nurses' communication, (3) discharge information, (4) communication about medication, (5) pain control, (6) physical environment of hospital, and (7) nursing assistance services. With the exception of item Q35, all items used in the scales exhibited factor loadings approximating or exceeding 0.40.

The internal consistency reliability coefficient was greater than 0.70 for six of the seven factors (Table 2). The scale "physical environment" was the exception with a Cronbach's α of 0.60. Except for item Q35, none of the Cronbach's α s (with the item deleted) exceeded the α for the total scale. The corrected item-total scale correlations ranged from 0.33 (item Q35: medical tests without pain) to 0.79 (item Q12: doctor listening carefully). On approximation only item Q35 did not meet the cut-off value of 0.40 for item-total scale correlations (Nunnally 1978).

Interitem correlations revealed that the similar item-pairs of Q46/47, Q49/50, and Q52/53 which detailed both written and verbal discharge information types found only in the Dutch instrument had moderate-to-low positive correlations of 0.40, 0.42, and 0.32, respectively. Interscale correlations ranged from 0.22 for nursing services and discharge information to 0.68 for nurses' communication and doctor's communication (Table 3). As the interscale correlations were all less than 0.70, the scales could be read as separate scales.

The seven scales were all significantly correlated with the three global ratings and the likelihood to recommend the hospital (Table 4). Nurses' communication exhibited the strongest significant association with nurses' global

Table 3: Interscale Correlations

	<i>Doctor's Care/ Communication</i>	<i>Nurses' Care/ Communication</i>	<i>Discharge Information</i>	<i>Medication</i>	<i>Pain Control</i>	<i>Physical Environment</i>	<i>Nursing Services</i>
Doctor's care/communication	1	0.68**	0.36**	0.48**	0.62**	0.44**	0.57**
Nurses' care/communication		1	0.25**	0.45**	0.63**	0.45**	0.67**
Discharge information			1	0.40**	0.34**	0.31**	0.22*
Medication				1	0.47**	0.35**	0.38**
Pain control					1	0.38	0.56**
Physical environment						1	0.44**
Nursing services							1

*Significant at $p < .05$;**Significant at $p < .01$.

Table 4: Correlations between Scales and Global Ratings

<i>Scale</i>	<i>Nurses' Global Rating</i>	<i>Doctor's Global Rating</i>	<i>Hospital Global Rating</i>	<i>Recommend Hospital</i>
Doctor's care/communication	0.54	0.76	0.60	0.57
Nurses' care/communication	0.72	0.55	0.62	0.59
Discharge information	0.24	0.34	0.32	0.34
Medication	0.42	0.38	0.36	0.39
Pain control	0.57	0.57	0.55	0.54
Physical environment	0.37	0.40	0.46	0.40
Nursing services	0.56	0.44	0.53	0.51

All correlations significant at $p < .01$.

The rankings/associations of the factors with the global ratings were also confirmed in multivariable regression analyses (not presented here).

rating ($r = 0.72$), hospital rating ($r = 0.62$), and the likelihood to recommend the hospital ($\rho = 0.59$), with all $p < .01$. Doctor's communication had the strongest correlation with the global rating of the doctor, but came in second on the global rating for and recommending the hospital. The least associated scale with any global rating or recommending hospital was discharge information, followed by communication about medication (except for nurses' rating).

Finally, as the survey instrument was aimed at providing information on the performance of each hospital corrected for their patients' characteristics, we ran multivariable regression models to determine the relationship between individual respondent characteristics and the global ratings (Table 5). We found that, overall, older age was significantly associated with higher ratings. Gender had mixed, but mostly nonsignificant, associations with the ratings by patients admitted for reasons other than childbirth (the models being corrected for age, education, general health status, and mental health status). Education also showed mixed results, higher education being insignificantly associated with lower ratings by surgical and medical patients, but was significantly associated with higher ratings for childbirth cases ($\beta = 0.17$, $p < .05$). Furthermore, poorer general health status was significantly related to lower global ratings by surgical (β ranging from -0.26 to -0.17 , $p < .001$ or $p < .05$) and medical patients (β ranging from -0.15 to -0.12 , $p < .05$). Mental health status exhibited no substantial associations with the global ratings given by any group of patients in the corrected models. Subsequently, only age and general health status consistently appeared to contribute to differences in between-hospital variations in patient experiences and ratings.

Table 5: Standardized Regression Coefficients of Patients' Characteristics on Their Global Ratings

<i>Patient Characteristics</i>	<i>Doctor's Global Rating</i>	<i>Nurses' Global Rating</i>	<i>Hospital Global Rating</i>
Surgical			
Age	0.22 ^{**¶}	0.15 ^{**¶}	0.18 ^{**¶}
Gender (female)	-0.01	0.02	-0.01
Some secondary education	-0.05	-0.05	-0.07
General health status*	-0.26 ^{**¶}	-0.17 [¶]	-0.25 ^{**¶}
Mental health status*	-0.04	-0.09	-0.09
R ²	0.11	0.06	0.12
Childbirth			
Age	0.18 [¶]	0.20 [¶]	0.26 [¶]
Some secondary education	0.06	0.17 [¶]	-0.017
General health status*	-0.11	-0.08	-0.15
Mental health status*	-0.09	-0.02	-0.01
R ²	0.02	0.04	0.04
Other medical			
Age	0.29 ^{**¶}	0.28 ^{**¶}	0.29 ^{**¶}
Gender (female)	-0.03	-0.004	0.01
Some secondary education	-0.04	-0.05	-0.08
General health status*	-0.14 [¶]	-0.12 [¶]	-0.15 [¶]
Mental health status*	-0.04	-0.08	0.001
R ²	0.09	0.08	0.10

All the models control for hospital effects.

*Health status self rated as excellent, very good, good, fair, or poor and were scored, respectively, from 1 to 5.

^{**} $p < .001$;

[¶] $p < .01$;

^{||} $p < .05$.

DISCUSSION

Patient experiences of care are an important data source for evaluating the functioning of the health care system (Cleary and Edgman-Levitan 1997; Cleary 1999). Like in the U.S., patient experience data are used for various purposes in the Netherlands. Recent policy reforms in the Netherlands encourage the use of experience data to provide patients/consumers and insurers information on which to base their choice and contracting of providers and institutions within a regulated health care market (Ministry of Health, Welfare and Sport 2002; Ministerie van Volksgezondheid, Welzijn en Sport 2004; ten Asbroek et al. 2004). Capturing inpatient care experiences augments the technical performance indicators being developed in the Dutch health care system. Moreover, Dutch insurers are looking to use valid performance

information from the patients' perspective to guide their hospital contracting, thus potentially widening the scope for the use of patient experience data. In this paper, we presented the initial psychometric properties of the pilot H-CAHPS in the Netherlands.

The items in the pilot instrument could be grouped into the following seven scales or composites: doctor's communication, nurses' communication, discharge information, communication about medication, pain control, physical environment of hospital, and nursing services. These are similar to the ones identified in the U.S. pilot H-CAHPS study (The CAHPS[®] II Investigators and the Agency for Healthcare Research and Quality 2003; CAHPS Survey Users Network 2004). The factor structure also reflects similar constructs as nursing care, doctor's care, information, hotel care, and discharge seen in existing Dutch instruments (Sixma et al. 1998; Hendriks et al. 2001, 2002; Jansen, Hutten, and Spreeuwenberg 2002). The factors reflect domains of hospital care experiences that patients may find important. Our results further suggest that these constructs may transcend cultural and contextual differences between Dutch and American patients in their understanding and interpretation of hospital care experiences. Additional studies including in-depth cognitive and confirmatory factor analyses will be needed to understand if this is really the case.

Items Q27 (family receiving help when visiting), Q28 (hospital staff introducing themselves), and Q43 (admission delay) had ambiguous loadings on the main factors, and this also appeared to be the case with the U.S. pilot results (The CAHPS[®] II Investigators and the Agency for Healthcare Research and Quality 2003), thus suggesting that they could be eliminated from the postpilot questionnaire and further analyses. Item Q35 (on medical tests without pain) is also problematic since it has low factor loadings, and did not improve the reliability coefficient of its corresponding scale.

The internal consistency reliability analysis showed that all but one of the seven factors had satisfactory coefficients greater than 0.70. The factor on physical environment had a low consistency coefficient of 0.60 as well as low corrected item-total correlations. This may be because of the low importance attached to this construct by the respondents, as shown in a previous Dutch study where patients were asked to rate which aspects of hospital performance they considered important (Jansen, Hutten, and Spreeuwenberg 2002). Therefore, given that many surveys have historically asked about amenities such as food, parking, and the physical environment, it may be useful to investigate critically the role that communication with providers plays in patient evaluations of hospital care.

Bivariate correlations between scales and between scales and global ratings supported the multifactor structure of the instrument, despite the moderately high correlations among doctor's and nurses' communication and nursing services. Moreover, they reflected the relatively high importance of nurses' and doctor's communication and pain management to patients' ratings of nurses, doctor, and hospital, and the likelihood to recommend the hospital. Discharge information, communication about medication, and the physical environment were less important in influencing patients' ratings. These findings suggest that scales closest to interpersonal care (Donabedian 1980) issues may trump the relative values of technical and environmental issues such as hospital comfort in the face of ill-health. Interestingly, like in the U.S. (The CAHPS[®] II Investigators and the Agency for Healthcare Research and Quality 2003) and in previous studies (Abramowitz, Cote, and Berry 1987; Rubin 1990), nurses' communication featured rather prominently in influencing global ratings and likelihood to recommend the hospital. This would be expected given the frequency and nature of nurse-patient interactions. Nevertheless, primary care research indicates that there may be no substantial differences in patient satisfaction with care provided by different types of practitioners (Roblin et al. 2004).

As this patient experience survey instrument was aimed at providing information on specific hospital's performance, it was important that the resulting information reflected the performance of the hospital, not differences because of their patients' characteristics. Our analysis suggested that age and self-reported general health status were among the most important patient characteristics to adjust for when reporting hospital performance (Hall and Dornan 1990; Cleary et al. 1992; Hall, Milburn, and Epstein 1993; Hargraves et al. 2001). Our analyses also supported the well-known trends in patient survey literature: higher ratings were seen among older patients, those with higher self-reported general health status (Hargraves et al. 2001) and the less educated (Hall and Dornan 1990), although the last group was not significant here. Female patients sometimes gave lower ratings (also not significant here) (Hargraves et al. 2001).

Although we have presented some good psychometric results of the new instrument to support its reliability and validity, several limitations existed in the study. First, there were only two hospitals involved in this study, hence limiting the potential for exploring hospital-level psychometric properties of the instrument. Second, for lack of resources, we were unable to conduct prior cognitive testing among patients to aid clarity, comprehension and establish equivalence. Third, because of Dutch privacy restrictions, we could not thoroughly analyze nonresponse in this study. Fortunately, the second

phase of the Dutch pilot is already underway to try to address most of these shortcomings using multiple hospital sites.

CONCLUSION

The H-CAHPS pilot instrument is reliable and valid for use within the Dutch context. Initial evaluation suggests that the American H-CAHPS constructs are not alien to the Dutch situation. In the future, it would be beneficial to pool the U.S. and Dutch data for further psychometric and performance analyses. This is necessary if equivalence and comparability are to be explored in-depth. This study contributes to the reform of the Dutch health care system aimed at increasing its transparency and performance.

ACKNOWLEDGMENTS

We thank the participants of the AcademyHealth's 2004 Annual Research Meeting held in San Diego, California, and of the International Society for Quality in Health Care's 2004 annual conference held in Amsterdam for their useful comments on earlier abstracts of this work. We also thank the RAND Graduate School in Santa Monica, California for their invitation to present and discuss the project on which this paper is based. We appreciate the support of Dr. Charles Darby (Agency for Healthcare Research and Quality). This study was funded by Agis Zorgverzekeringen, the Netherlands. We are also grateful to the editors and two anonymous reviewers of the Journal, *Health Services Research*, for their helpful comments.

NOTE

1. This is addressed in "Delnoij DMJ et al. Made in the USA: the import of the American Consumer Assessment of Health Plans Study surveys into the Dutch insurance system (*submitted manuscript*)."

REFERENCES

- Abramowitz, S., A. A. Cote, and E. Berry. 1987. "Analyzing Patient Satisfaction: A Multianalytic Approach." *Quality Review Bulletin* 13 (4): 122-30.
- Arah, O. A., N. S. Klazinga, D. M. J. Delnoij, A. H. A. ten Asbroek, and T. Custers. 2003. "Conceptual Frameworks for Health Systems Performance: A Quest for

- Effectiveness, Quality and Improvement.” *International Journal for Quality in Health Care* 15 (5): 377–98.
- CAHPS Survey Users Network. 2004. “Update on Hospital CAHPS.” *The CAHPS[®] Connection* 1 (May 2004): 2–3.
- CAHPS[®] Survey Users Network (SUN). 2004. “An Introduction to CAHPS” [accessed November 30, 2004]. Available at <http://www.cahps-sun.org/Home/Introduction.asp>
- Carey, R. G., and J. H. Seibert. 1993. “A Patient Survey System to Measure Quality Improvement: Questionnaire Reliability and Validity.” *Medical Care* 31 (9): 834–45.
- Centers for Medicare and Medicaid Services. 2004. “Hospital Quality Initiative” [accessed November 30, 2004]. Available at <http://www.cms.hhs.gov/quality/hospital/>
- Cleary, P. D. 1999. “The Increasing Importance of Patient Surveys.” *British Medical Journal* 319: 720–1.
- Cleary, P. D., and S. Edgman-Levitan. 1997. “Health Care Quality. Incorporating Consumer Perspectives.” *Journal of the American Medical Association* 278 (19): 1608–12.
- Cleary, P. D., S. Edgman-Levitan, W. McMullen, and T. L. Delbanco. 1992. “The Relationship between Reported Problems and Patient Summary Evaluations of Hospital Care.” *Quality Review Bulletin* 18 (2): 53–9.
- Cronbach, L. J. 1951. “Coefficient Alpha and the Internal Structure of Tests.” *Psychometrika* 16: 297–334.
- Donabedian, A. 1980. *Explorations in Quality Assessment and Monitoring. Vol. 1. The Definition of Quality and Approaches to its Assessment*. Ann Arbor, MI: Health Administration Press.
- Farley, D. O., P. F. Short, M. N. Elliott, D. E. Kanouse, J. A. Brown, and R. D. Hays. 2002. “Effects of CAHPS Health Plan Performance Information on Plan Choices by New Jersey Medicaid Beneficiaries.” *Health Services Research* 37 (4): 985–1007.
- Goldstein, E., and J. Fyock. 2001. “Reporting of CAHPS Quality Information to Medicare Beneficiaries.” *Health Services Research* 36 (3): 477–88.
- Hall, J. A., and M. C. Dornan. 1990. “Patient Sociodemographic Characteristics as Predictors of Satisfaction with Medical Care: A Meta-Analysis.” *Social Science and Medicine* 30 (7): 811–8.
- Hall, J. A., M. A. Milburn, and A. M. Epstein. 1993. “A Causal Model of Health Status and Satisfaction with Medical Care.” *Medical Care* 31 (1): 84–94.
- Hargraves, J. L., R. D. Hays, and P. D. Cleary. 2003. “Psychometric Properties of the Consumer Assessment of Health Plans Study (CAHPS) 2.0 Adult Core Survey.” *Health Services Research* 38 (6, part 1): 1509–27.
- Hargraves, J. L., I. B. Wilson, A. Zaslavsky, C. James, J. D. Walker, G. Rogers, and P. D. Cleary. 2001. “Adjusting for Patient Characteristics When Analyzing Reports from Patients about Hospital Care.” *Medical Care* 39 (6): 635–41.
- Hays, R. D., J. A. Shaul, V. S. Williams, J. S. Lubalin, L. D. Harris-Kojetin, S. F. Sweeny, and P. D. Cleary. 1999. “Psychometric Properties of the CAHPS 1.0 Survey Measures.” *Medical Care* 37 (3 suppl): MS22–31.

- Hekking, C. F., H. J. Sixma, L. Wigersma, C. J. Yzermans, J. T. Van Der Meer, P. J. Bindels, K. Brinkman, and S. A. Danner. 2003. "QUOTE-HIV: An Instrument for Assessing Quality of HIV Care from the Patients' Perspective." *Quality and Safety in Health Care* 12 (3): 188–93.
- Hendriks, A. A., F. J. Oort, M. R. Vrieling, and E. M. Smets. 2002. "Reliability and Validity of the Satisfaction with Hospital Care Questionnaire." *International Journal for Quality in Health Care* 14 (6): 471–82.
- Hendriks, A. A., M. R. Vrieling, S. Q. Es, and E. M. Smets. 2000. *Patiënten-Tevredenheids Onderzoek: Eindrapport*. Amsterdam: Academisch Medisch Centrum, Universiteit van Amsterdam.
- Hendriks, A. A., M. R. Vrieling, E. M. Smets, S. Q. van Es, and J. C. De Haes. 2001. "Improving the Assessment of (in)Patients' Satisfaction with Hospital Care." *Medical Care* 39 (3): 270–83.
- Iezzoni, L. I. 2002. "Toward Universal Design in Assessing Health Care Experiences." *Medical Care* 40 (9): 725–8.
- Jansen, A. F. C., J. B. F. Hutten, and O. Spreeuwenberg. 2002. *Patiënten Oordelen over Ziekenhuizen II. Ervaringen en Meningen van Patiënten in 37 Nederlandse Ziekenhuizen*. Utrecht: Nivel and Consumentenbond.
- Ministerie van Volksgezondheid, Welzijn en Sport. 2004. *Beleidsagenda 2005. Nieuwe Fundamenten Voor Een Toekomstbestendige Zorg*. Den Haag: Ministerie van Volksgezondheid, Welzijn en Sport.
- Ministry of Health, Welfare and Sport. 2002. *A Question of Demand. Outlines of the Reform of the Health Care System in the Netherlands*. The Hague: Ministry of Health, Welfare and Sport (International Publication Series Health, Welfare and Sport no. 14E).
- Morales, L. S., M. N. Elliott, R. Weech-Maldonado, K. L. Spritzer, and R. D. Hays. 2001. "Differences in CAHPS Adult Survey Reports and Ratings by Race and Ethnicity: An Analysis of the National CAHPS Benchmarking Data 1.0." *Health Services Research* 36 (3): 595–617.
- Nunnally Jr., J. C. 1978. *Psychometric Theory*, 2d Edition. New York: McGraw-Hill.
- Roblin, D. W., E. R. Becker, E. K. Adams, D. H. Howard, and M. H. Roberts. 2004. "Patient Satisfaction with Primary Care: Does Type of Practitioner Matter?" *Medical Care* 42 (6): 579–90.
- Rubin, H. R. 1990. "Can Patients Evaluate the Quality of Hospital Care?" *Medical Care Review* 47 (3): 267–326.
- Sixma, H. J., J. J. Kerssens, C. V. Campen, and L. Peters. 1998. "Quality of Care from the Patients' Perspective: From Theoretical Concept to a New Measuring Instrument." *Health Expectations* 1 (2): 82–95.
- Sixma, H. J., P. M. Spreeuwenberg, and M. A. van der Pasch. 1998. "Patient Satisfaction with the General Practitioner: A Two-Level Analysis." *Medical Care* 36 (2): 212–29.
- Sixma, H. J., C. van Campen, J. J. Kerssens, and L. Peters. 2000. "Quality of Care from the Perspective of Elderly People: The QUOTE-Elderly Instrument." *Age and Ageing* 29 (2): 173–8.

- Solomon, L. S., R. D. Hays, A. M. Zaslavsky, L. Ding, and P. D. Cleary. 2005. "Psychometric Properties of a Group-Level Consumer Assessment of Health Plans Study (CAHPS) Instrument." *Medical Care* 43 (1): 53–60.
- Spranca, M., D. E. Kanouse, M. Elliott, P. F. Short, D. O. Farley, and R. D. Hays. 2000. "Do Consumer Reports of Health Plan Quality Affect Health Plan Selection?" *Health Services Research* 35 (5, part 1): 933–47.
- SPSS Inc. 2004. *Statistical Package for the Social Sciences. SPSS (version 12.0.2)*. Chicago: SPSS Inc.
- Stevens, J. P. 1992. *Applied Multivariate Statistics for Social Sciences*, 2d Edition. Hillsdale, NJ: Erlbaum.
- Streiner, D. L., and G. R. Norman. 2003. *Health Measurement Scales: A Practical Guide to Their Development and Use*, 3d Edition. Oxford: Oxford University Press.
- ten Asbroek, A. H. A., O. A. Arah, J. Geelhoed, T. Custers, D. M. J. Delnoij, and N. S. Klazinga. 2004. "Developing a National Performance Indicator Framework for the Dutch Health System." *International Journal for Quality in Health Care* 16 (suppl 1): i65–71.
- The CAHPS[®] II Investigators and the Agency for Healthcare Research and Quality. 2003. *HCAHPS Three-State Pilot Study Analysis Results*. Baltimore: Centers for Medicare and Medicaid Services (Available at http://www.cms.hhs.gov/quality/hospital/3State_Pilot_Analysis_Final.pdf).
- Weidmer, B., J. Brown, and L. Garcia. 1999. "Translating the CAHPS 1.0 Survey Instruments into Spanish." *Medical Care* 37 (3 suppl): MS89–96.

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article online:

APPENDIX S1. Pilot Core Items and Differences between the U.S. and Dutch Versions