**Validity and reliability of the HeartQoL questionnaire in a large sample of stable coronary patients: The EUROASPIRE IV Study.**

Delphine De Smedt[1], Els Clays[1], Stefan Höfer[2], Neil Oldridge[3], Kornelia Kotseva[4], Dirk De Bacquer[1] on behalf of the EUROASPIRE Investigators

[1]Department of Public Health, Ghent University, Gent, Belgium

[2]Department of Medical Psychology, Innsbruck Medical University, Innsbruck, Austria

[3]College of Health Sciences, University of Wisconsin-Milwaukee, Wisconsin, USA

[4]Department of Cardiovascular Medicine, National Heart and Lung Institute, Imperial College London, London, UK

***Abstract:***

Background: The aim of this study was to investigate the validity and reliability of the HeartQoL instrument in a population of stable coronary patients.

Design: Analyses are based on the cross-sectional EUROASPIRE IV (EUROpean Action on Secondary and Primary prevention through Intervention to Reduce Events) survey.

Methods: Patients with a diagnosis of coronary heart disease were examined and interviewed 6 months to 3 years after their coronary event. During the interview patients were asked to fill out the HeartQoL health-related quality of life questionnaire as well as the EuroQoL-5D and the Hospital Anxiety and Depression Scale questionnaires. Psychometric analyses assessing the reliability and validity of the HeartQoL instrument were performed.

Results: A total of 7449 patients completed the HeartQoL instrument. Cronbach's alpha indicated excellent internal consistency for the global HeartQoL scale ($\alpha$=0.92) and the physical subscale ($\alpha$=0.91) and good internal consistency for the emotional subscale ($\alpha$=0.87). Factor analysis confirmed the two-dimensional construct although mixed results were found regarding the model fit. Discriminative validity analysis confirmed better HeartQoL results in males, younger patients and higher educated persons. Likewise convergent validity was confirmed with moderate to strong correlations among hypothesized constructs.

Conclusion: Overall, psychometric analyses of the HeartQoL instrument in a population of patients with stable coronary heart disease showed good reliability and validity both at the European as well as the country-specific level. However, further research should focus on the responsiveness of the HeartQoL, the ceiling effect of the emotional subscale, construct validity, and the clinically meaningful difference.

***Introduction***

Life expectancy has increased substantially in Europe from 65.6 years in the late 1950's to 76.5 years currently (1). While the drop in coronary heart disease (CHD) mortality is partly responsible for this evolution, population ageing is also associated with an increase in CHD morbidity, accounting for 10% of total disability adjusted life years lost (2). In addition to the focus on clinical measures and functional outcomes, an increased awareness of the importance of patient-reported outcomes such as health-related quality of life (HRQL) has emerged (3). Within this context, several tools have been developed to assess patient's physical, emotional and social well-being. Many generic instruments such as the EuroQoL-5D and the short form 36 exist, allowing comparison across different patient groups (4-6). Furthermore, a number of disease-specific instruments are available, useful in specific patient groups and often more sensitive in capturing small differences between patients, e.g Seattle Angina Questionnaire (SAQ) for patients with angina (7) and the MacNew Heart Disease quality of life questionnaire (MacNew) originally designed for patients with myocardial infarction (MI) but also valid in patients with angina and heart failure (8). Since comparison between the different patients groups is difficult with the existing disease-specific tools, the HeartQoL questionnaire was designed to be a single reliable and valid core HRQL instrument in patients with CHD for making between-diagnosis comparisons possible and to assess the change in HRQL after treatment (9;10). The HeartQoL consists of 14 items; 10 items focus on physical well-being and 4 items on emotional well-being, together providing a global scale.

The psychometric properties of the HeartQoL have been assessed in the original HeartQoL Project sample of 6384 patients from 22 countries being actively treated for one of the three major CHD diagnoses, i.e., angina, MI, and ischaemic heart failure (9). The aim of the current study was to determine and confirm or extend (as some languages were not available in the original study) the reliability and validity of the HeartQoL in an independent large European sample of 7449 patients with

CHD from 24 different European countries who were in a stable phase of their disease using the EUROASPIRE IV (EUROpean Action on Secondary and Primary Prevention through Intervention to Reduce Events) survey data (11).

***Methods***

*Study population and data collection*

Analyses were based on data gathered during the recent EUROASPIRE IV survey, performed in 2012-2013 in 24 European countries: Belgium, Bosnia Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Finland, France, Germany, Greece, Ireland, Latvia, Lithuania, Netherlands, Poland, Romania, Russian Federation, Serbia, Slovenia, Spain, Sweden, Turkey, Ukraine and United Kingdom (11). Patients eligible for inclusion were men or women, aged ≥18 years and <80 years, who had been hospitalized for a first or recurrent coronary event, defined as the recruiting event, including elective or emergency coronary artery bypass graft surgery (CABG); elective or emergency percutaneous coronary intervention (PCI); first or recurrent acute ST elevation and non-ST elevation myocardial infarction (MI); or acute myocardial ischemia. Patients were interviewed 6 months to 3 years following the recruiting event.

Patient reported outcomes were assessed using 3 questionnaires: the HeartQoL (10), the EuroQol-5D-5L (EQ-5D-5L) (12), and the Hospital Anxiety and Depression Scale (HADS) (13). In each country, questionnaires were administered in the official language.

The HeartQoL comprises 14 items with 4 response categories ranging from “not bothered” to ”bothered a lot”. In the original HeartQoL development paper by Oldridge et al. (2014), a bi-dimensional instrument was proposed with a 10-item physical subscale and a 4-item emotional subscale (10). Both the global scale (all items), and the physical (10 items) and emotional (4 items) subscale scores, ranging between 0 (poor HRQL) and 3 (better HRQL), can be calculated as the mean of the scored items. Validated language

versions of the HeartQoL questionnaire were not yet available for some countries included in the EUROASPIRE study (Bulgaria, Bosnia & Herzegovina, Czech Republic, Croatia, Cyprus, Finland, Greece, Latvia, Lithuania, Romania, Serbia, Slovenia, Turkey), and a forward and backward translation methodology was used in their translation (14).

EQ-5D-5L is a widely used simple, generic health utility instrument to measure health status in a standardized way (4;15;16). The measure is applicable to a wide range of conditions and treatments and contains a self-classifier made up of 5 Likert-scale items on mobility, self-care, usual activities, pain/discomfort, and anxiety/depression from which a single summary index can be derived (17) and a visual analogue scale asking the patient to indicate on a 20cm vertical scale how good or bad his/her health status is today. Within this study, the index scores were based on UK value set for all countries (18). Index values can range between -0.594 and 1 with 1 representing perfect health, 0 representing death, and <0 representing a health state perceived worse than death.

The HADS instrument, containing 14 items with four response categories each, permits calculation of both an anxiety score (HADS-A) and a depression score (HADS-D). The total score on each subscale ranges between 0 and 21. A score < 8 can be considered as being in the normal range, higher scores indicate a possible or probable disorder (13;19).

*Psychometric analyses*

Reliability and validity of the HeartQoL instrument were assessed in the current study (20). Stratification by country, gender, age group (<60 years, 60-69 years and ≥70 years) and educational level (primary school or less; secondary school completed, high school completed or intermediate between secondary level completed; university/college degree or equivalent) was performed. A conservative approach was applied, excluding patient records from the analyses for which at least 1 HeartQoL item was missing. All

analyses were performed using IBM SPSS 22, except the exploratory (EFA) and confirmatory factor analyses (CFA), which were conducted using M-plus version 7.2.

Reliability

The Cronbach's alpha was used to assess the internal consistency of the instruments. A threshold of 0.7 was considered acceptable, a value >0.8 was considered as good, and a value >0.9 as excellent (21).

Validity

EFA was performed on our sample in order to explore the underlying factor structure. In addition, CFA was used to test whether the original scale construct, proposed by Oldridge et al. (2014), could be confirmed in our sample (10). The Weighted Least Squares Means and Variance adjusted (WLSMV) estimation method was applied with ordinal data. Multiple goodness of fit tests were used to assess the model fit, including the Root mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI). The CFI and TLI range between 0 and 1 with a value >0.9 being considered as an acceptable model fit (22). RMSEA values also range between 0 and 1 with a value <0.06 being considered as a good fit, a value <0.08 was considered as an acceptable fit and a value >0.1 led to rejection of the model (23;24). Factor loadings represent the correlation between items and factor. Standardized factor loadings >0.5 were perceived as good, loadings >0.4 indicated an acceptable correlation and those <0.4 were perceived as not good.

Discriminative validity was examined with the known-groups method based on multilevel linear models to account for the clustering of patients within countries (25). Based on literature findings we hypothesized that HRQL would decrease with age and lower education, and would be lower in females and anxious or depressive patients (26-30).

Convergent validity was evaluated to assess the correlations between theoretically correlated constructs using Spearman correlation coefficients due to the non-normal distribution of the constructs.

Correlations <0.3 were considered as negligible, a value between 0.3 and 0.5 as moderate and >0.5 as strong (31). We hypothesized that the global HeartQoL scale and both the HeartQoL physical and the HeartQoL emotional subscales would correlate strongly with the EQ-$5D_{index}$. Furthermore the HeartQoL emotional was expected to correlate well with HADS.

In line with the paper by Oldridge et al. on the reliability and validity of the tool in the original sample, we assessed the ceiling and floor effect of the tool in comparison with the EQ-5D and the HADS tools (9).

***Results***

*Patient characteristics*

From the 16,426 medical records from patients eligible for inclusion, 7998 patients (48.7%) were interviewed and examined. The HeartQoL instrument was completed by 7449 patients (93.1%) whose mean age was 64 years (ranging from 58.5 to 67.3 years between countries). The majority of patients (76.1%) were male (ranging from 61.6% to 90.5% between countries). The distribution of the recruiting event was as follows: 13.3% CABG, 55.9% PCI, 21.1% AMI, 9.7% ischaemia. Education levels were as follows: primary education, 16.6%; secondary education, 60.6%; higher education, 22.8%.

*Health-related quality of life*

The overall and country-specific mean values on the different HRQL constructs are shown in Table 1. The mean HeartQoL (±SD) scores were as follows: global scale, 2.18±0.66; physical subscale, 2.13±0.72 and emotional subscale, 2.30±0.72. The HeartQoL ceiling effects ranged from 8.1% (global scale) to 28.7% (emotional subscale); the floor effects were always less than 1.2% (Table 1).

*Reliability*

Overall, reliability was good to excellent on the global scale and both the physical and emotional subscales. Some variation was seen between countries, but for all countries, a good to excellent internal consistency was observed on the global scale (0.85-0.95) and the physical subscale (0.83-0.95) except for Bosnia Herzegovina (0.65 and 0.44 resp.) (Table 2). For the HeartQoL emotional subscale, the overall Cronbach's alpha was 0.87, ranging between 0.80 and 0.93 except for Cyprus (0.79), Bosnia Herzegovina (0.77) and Poland (0.67).

*Validity*

*Conceptual model*

On the other hand, mixed results were found regarding the construct validity of the tool. Using the data from all patients, the two-factor structure was proposed with EFA. For one item (Q8), EFA factor loadings were similar on both constructs. With CFA, both the TLI (0.94) and the CFI (0.95) showed good fit with the originally proposed constructs, whereas the RMSEA (0.117) suggested rejecting the model. Excellent standardized factor loadings >0.70 were found with the CFA, representing a very good correlation between observed variables and extracted components. We explored the data further in an attempt to resolve the contradictory results among the fit indices. Replacing Q8 from the physical component to the emotional component did not alter the findings and the results were not improved by loading Q8 on both factors. Similarly, performing the CFA only in those countries that participated in the original validation study (Belgium, France, Germany, Hungary, Ireland, Italy, Netherlands, Poland, Russia, Spain, Sweden, Ukraine, UK) did not alter the results. Finally, performing the CFA analyses on a country-specific level gave similar results for all countries with the exception of Bosnia for which a worse CFA outcome was seen.

*Discriminative validity*

There were gender, age, and educational level differences in the global, physical, and emotional scores (all p<0.001) (Table 4). Females, anxious and depressed patients and primary educated patients reported poorer global, physical, and emotional HeartQoL scores. Older patients reported poorer global and physical HeartQoL scores whereas younger patients reported poorer HRQL on the emotional subscale.

*Convergent validity*

Convergent validity was demonstrated with the HeartQoL global scale, and both the physical and the emotional subscales correlating strongly with the EQ-$5D_{index}$ and moderately with the EQ-VAS. Furthermore, the global HeartQoL and emotional subscale correlated strongly with both HADS scales (table 5).

*Reported problems*

In patients with full health on the EQ-5D (no problems on all items), 25.3% were not bothered at all according to the global HeartQoL score (mean=2.70±0.38), 32.0% were not bothered by any physical subscale item (mean=2.67±0.44), and 58.6% were not bothered by any emotional subscale item (mean=2.76±0.42). Of all the patients indicating best imaginable health status on the EQ-VAS, 31.5% were not bothered at all according to the global scale (mean=2.62±0.54), 37.4% were not bothered on the physical subscale (mean=2.60±0.57) and 62.6% were not bothered according to the emotional subscale (mean=2.66 ±0.60).

***Discussion***

We evaluated the reliability and validity of the HeartQoL questionnaire based on data observed in the EUROASPIRE IV survey on a European sample of 7998 stable coronary patients. Overall, the HeartQoL instrument showed to be a reliable tool for measuring HRQL in patients with stable CHD with excellent internal consistency on the global HeartQoL scale and physical subscale and good internal consistency on

the emotional subscale. Although EFA proposed the two-factor structure, mixed results were found with CFA regarding the construct validity of the tool. Convergent validity was supported by moderate to strong correlations between theoretically correlated constructs and discriminative validity was fully confirmed with highly significant group differences on all scales.

The internal consistency results in the present study were similar to the initial reliability observations by Oldridge et al. who reported Cronbach's alpha for the global, physical and emotional HeartQoL of 0.91, 0.90 and 0.81, respectively (9). Performing the reliability analyses on a country-specific level gave similar results for all countries with the exception of Bosnia Herzegovina.

On the other hand, mixed results were found regarding the construct validity of the tool. Using the complete sample, the two-factor structure was proposed with EFA. For one item (Q8), factor loadings were similar on both constructs. With CFA, both the TLI and the CFI showed good fit with the originally proposed constructs, whereas the RMSEA suggested rejecting the model. Further exploration of the data was done in an attempt to resolve this issue. Replacing Q8 from the physical component to the emotional component did not alter the results and the results were not improved by loading Q8 on both factors. Similarly, performing the CFA only in those countries that participated in the original validation study did not alter the findings. Finally, performing the CFA analyses on a country-specific level gave similar results for all countries with the exception of Bosnia for which a worse CFA outcome was seen.

Discriminative validity of the global HeartQoL scale and both subscales was confirmed with better HeartQoL scores seen in male patients, younger patients, patients with a higher educational level and patients without HADS-A or HADS-D symptoms. These results are supported by literature findings reporting on other HRQL tools (26-30). Convergent validity was supported with moderate to strong correlations between theoretically correlated constructs.

The ceiling (best possible score) and floor effect (worst possible score) seen in the EUROASPIRE patient population was similar to the original HeartQoL study (9). Floor and ceiling effects of 1% to 15% are considered small (32). The absence of floor effects indicates that the tool is sensitive for assessing adverse disease progression. In this respect, the HeartQoL tool performs better than the HADS instrument and similar to the EQ-5D instrument. Furthermore, the ceiling effect of the global and physical HeartQoL scales also showed good results. Not unexpected - as it is based on only 4 items – a substantial ceiling effect on the emotional subscale was seen.

Test-retest and responsiveness could not be tested due to the cross-sectional design of the survey. Interpretation of scores needs to be further investigated. A small number of patients, 7% of the total patient group, did not provide HeartQoL information; as this subgroup had a slightly worse risk profile (more male, more ischemia, lower education, worse EQ-5D and HADS), this potentially could have resulted in a small bias. Furthermore, for some countries (Bulgaria, Bosnia & Herzegovia, Czech Republic, Croatia, Cyprus, Finland, Greece, Latvia, Lithuania, Romania, Serbia, Slovenia, Turkey) a validated version of the HeartQoL was not yet available, and even though a standardized forward-backward translation approach was used, some translation bias might have occurred.

In summary, the HeartQoL questionnaire showed adequate reliability and validity for measuring HRQL in patients with stable CHD in most languages/countries. However, further research should focus on the ceiling effect of the emotional component and its sensitivity to capture small changes, as well as on construct validity, and the clinically meaningful difference.

**Acknowledgment**

The authors thank the administrative staff, physicians, nurses, and other personnel at the hospitals in which the study was carried out, and all the patients who participated in the EUROASPIRE studies. The EUROASPIRE IV survey was carried out under the auspices of the European Society of Cardiology, EURObservational Research Programme. Unrestricted educational grants to the European Society of Cardiology were obtained from Amgen, AstraZeneca, Bristol-Myers Squibb and AstraZeneca, F. Hoffman-La Roche, GlaxoSmithKline, and Merck Sharp & Dohme. The sponsors of the EUROASPIRE surveys had no role in the design, data collection, data analysis, data interpretation, decision to publish, or writing the manuscript. Delphine De Smedt is financially supported by the Research Foundation Flanders.

**References**


(1) United Nations. Life Expectancy at Birth.Available from: http://data.un.org/Data.aspx?d=PopDiv&f=variableID%3A68

(2) Nichols M, Townsend N, Scarborough P, et al. European Cardiovascular Disease Statistics. 2012.

(3) Anker SD, Agewall S, Borggrefe M, et al. The importance of patient-reported outcomes: a call for their comprehensive integration in cardiovascular clinical trials. Eur Heart J 2014;35(30):2001-9.

(4) Bowling A. Measuring Disease. A Review of Disease-specific Quality of Life Measurement Scales. Buckingham: Open University Press, 2001.

(5) Ferrans CE, Zerwic JJ, Wilbur JE, Larson JL. Conceptual model of health-related quality of life. J Nurs Scholarsh 2005;37(4):336-42.

(6) Ware JE, Jr. SF-36 health survey update. Spine 2000;25(24):3130-9.

(7) Spertus JA, Winder JA, Dewhurst TA, et al. Development and evaluation of the Seattle Angina Questionnaire: a new functional status measure for coronary artery disease. Journal of the American College of Cardiology 1995;25(2):333-41.

(8) Hofer S, Saleem A, Stone J, et al. The MacNew Heart Disease Health-Related Quality of Life Questionnaire in patients with angina and patients with ischemic heart failure. Value Health 2012;15(1):143-50.

(9) Oldridge N, Höfer S, McGee H, et al. The HeartQoL: part II. Validation of a new core health-related quality of life questionnaire for patients with ischemic heart disease. Eur J Prev Cardiol 2014;21(1):98-106.

(10) Oldridge N, Höfer S, McGee H, et al. The HeartQoL: Part I. Development of a new core health-related quality of life questionnaire for patients with ischemic heart disease. Eur J Prev Cardiol 2014;21(1):90-7.

(11) Kotseva K, Wood D, De Bacquer D, et al. EUROASPIRE IV: A European Society of Cardiology survey on the lifestyle, risk factor and therapeutic management of coronary patients from twenty four European countries. European Journal of Preventive Cardiology 2015;DOI10.1177/2047487315569401.

(12) Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011;20(10):1727-36.

(13) Zigmond A, Snaith R. The hospital anxiety and depression scale. Acta Psychiatrica Scandinavica 1983;67(6):361-70.

(14) Acquadro C, Conway K, Hareendran A, Aaronson N. Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. Value Health 2008;11(3):509-21.

(15) Dolan P. Modeling valuations for EuroQol health states. Medical Care 1997;35(11):1095-108.

(16) Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Annals of Medicine 2001;33(5):337-43.

(17) Oemar M, Janssen B. EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument. 2013.

(18) van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health 2012;15(5):708-15.

(19) Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002;52(2):69-77.

(20) Hays RD, Anderson R, Revicki D. Psychometric considerations in evaluating health-related quality of life measures. Quality of Life Research 1993;2(6):441-9.

(21) Nunnally JC, Bernstein IH. Psychometric theory (3rd ed ed.). New York: McGraw-Hill, 1994.

(22) Bentler P, Bonett D. Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin 1980;88:588-606.

(23) Brown M, Cudeck R. Alternative ways of assessing model fit. Sociological Methods and Research 1992;21:-230.

(24) Hu L, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: Coventional criteria versus new alternatives. Structural Equation Modeling 1999;6(1):1-55.

(25) Vandereyt F, Dendale P, Vanhees L, et al. Psychometric properties of the Flemish version of the MacNew heart disease health-related quality of life questionnaire. Acta Cardiol 2012;67(1):31-9.

(26) De Smedt D, Clays E, Doyle F, et al. Validity and reliability of three commonly used quality of life measures in a large European population of coronary heart disease patients. Int J Cardiol 2013;167(5):2294-9.

(27) Leung Yinko SS, Pelletier R, Behlouli H, et al. Health-related quality of life in premature acute coronary syndrome: does patient sex or gender really matter? J Am Heart Assoc 2014;3(4).

(28) Maes S, de Gucht V, Goud R, et al. Is the MacNew quality of life questionnaire a useful diagnostic and evaluation instrument for cardiac rehabilitation? Eur J Cardiovasc Prev Rehabil 2008;15(5):516-20.

(29) Norris CM, Murray JW, Triplett LS, Hegadoren KM. Gender roles in persistent sex differences in health-related quality-of-life outcomes of patients with coronary artery disease. Gend Med 2010;7(4):330-9.

(30) Xie J, Wu E, Zheng Z, et al. Patient-reported health status in coronary heart disease in the United States: age, sex, racial, and ethnic differences. Circulation 2008;118(5):491-7.

(31) Cohen J. Statistical power analysis for the behavioral sciences (2nd ed ed.). Hillsdale, N.J: L. Erlbaum Associates, 1988.

(32) McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995;4(4):293-307.