ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

Patient-Reported Morbidity Instruments: A Systematic Review

Arvind Oemrawsingh, MD, MHS,[1,*] Nishwant Swami, BA,[2] José M. Valderas, MD, PhD, MPH,[3] Jan A. Hazelzet, MD, PhD,[1] Andrea L. Pusic, MD, MHS, FACS, FRCSC,[4] Richard E. Gliklich, MD,[5] Regan W. Bergmark, MD[5,6]

[1]Department of Public Health, Erasmus University Medical Center, Rotterdam, The Netherlands; [2]University of Massachusetts Medical School, Worcester, MA, USA; [3]International Society for Quality of Life Research (ISOQOL), Health Services & Policy Research, University of Exeter Medical School; Exeter, England, UK; [4]Division of Plastic and Reconstructive Surgery, Patient Reported Outcomes, Value, and Experience (PROVE) Center, Brigham and Women's Hospital, Boston, MA, USA; [5]Department of Otolaryngology – Head and Neck Surgery, Harvard Medical School, Boston, MA, USA; [6]Center for Surgery and Public Health, Patient Reported Outcomes, Value and Experience (PROVE) Center, Brigham and Women's Hospital, Boston, MA, USA.

ABSTRACT

Objectives: Although comorbidities play an essential role in risk adjustment and outcomes measurement, there is little consensus regarding the best source of this data. The aim of this study was to identify general patient-reported morbidity instruments and their measurement properties.

Methods: A systematic review was conducted using multiple electronic databases (Embase, Medline, Cochrane Central, and Web of Science) from inception to March 2018. Articles focusing primarily on the development or subsequent validation of a patient-reported morbidity instrument were included. After including relevant articles, the measurement properties of each morbidity instrument were extracted by 2 investigators for narrative synthesis.

Results: A total of 1005 articles were screened, of which 34 eligible articles were ultimately included. The most widely assessed instruments were the Self-Reported Charlson Comorbidity Index (n = 7), the Self-Administered Comorbidity Questionnaire (n = 3), and the Disease Burden Morbidity Assessment (n = 3). The most commonly included conditions were diabetes, hypertension, and myocardial infarction. Studies demonstrated substantial variability in item-level reliability versus the gold standard medical record review (κ range 0.66-0.86), meaning that the accuracy of the self-reported comorbidity data is dependent on the selected morbidity.

Conclusions: The Self-Reported Charlson Comorbidity Index and the Self-Administered Comorbidity Questionnaire were the most frequently cited instruments. Significant variability was observed in reliability per comorbid condition of patient-reported morbidity questionnaires. Further research is needed to determine whether patient-reported morbidity data should be used to bolster medical records data or serve as a stand-alone entity when risk adjusting observational outcomes data.

Keywords: comorbidity, health services, morbidity, patient report, psychometrics, self-report, surveys and questionnaires.

VALUE HEALTH. 2020; ■(■):■–■

Introduction

Value-based healthcare (VBHC) initiatives rely on risk adjustment to compare patient populations across hospitals. In addition to understanding the index disease of interest, comorbid conditions are necessary for case-mix adjustment. "Morbidity" is defined here as the presence of medical conditions. Clinicians are increasingly grappling with the challenges of treating patients with multiple co-occurring diseases (multimorbidity). In addition to treatment difficulties, multimorbidity is often associated with worse outcomes including decreased quality of life, psychological distress, longer hospital stays, more postoperative complications, higher cost of care, and higher mortality.[1,2]

To identify opportunities for outcomes improvement, registries and groups like the International Consortium for Healthcare Outcomes Measurement (ICHOM) have attempted to standardize and compare observational data across hospitals.[3] These comparative studies often rely on risk adjustment algorithms to account for clinical differences in patient populations.[4] In analyzing a changing medical landscape with more

Conflict of interest: Swami was a paid employee at the International Consortium for Health Outcomes Measurement (ICHOM) during the development of this paper. Pusic reports personal fees from Patient Reported Outcome Measures (Q-PROMS) outside the submitted work.

* Address correspondence to: Arvind Oemrawsingh, MD, MHS, Department of Public Health, Erasmus Medical Center, Room Na-2403, P.O. Box 2040, 3000 CA Rotterdam, the Netherlands. Email: a.oemrawsingh@erasmusmc.nl

https://doi.org/10.1016/j.jval.2020.02.006

multimorbidity patients, reliable morbidity data has increasingly become a focal point for fair benchmarking as part of the shift toward VBHC.

Accurate inclusion of comorbidities in large data sets has proven to be a vexing problem. Although morbidity plays a crucial role in risk adjustment, risk stratification, and outcomes measurement, there is little consensus regarding the best source of this data. Comparisons of morbidity data from different sources have displayed significant variations.[5] Notable inconsistencies have been observed when morbidity data is collected from administrative sources, such as claims data.[6] Administrative data generally underreports comorbid conditions, leading to a lack of accounting for overall level of sickness of the patient.[7-9] Although some studies have shown more accurate information in hospital chart reviews, concerns arise regarding the burden of collection and the feasibility of wide-scale use.[10,11]

To obtain more accurate morbidity data feasibly, clinicians have increasingly turned to patient-reported instruments as a potential alternative.[12,13] The objective of this study was to provide a comprehensive evidence base of validated patient-reported morbidity instruments to aid in the selection of these instruments for use in clinical practice. Although many disease-specific morbidity instruments exist, our study examined questionnaires applicable to the broader patient population to allow for broader implementation across a healthcare system.

Methods

Design and Rationale

Risk adjustment for the comparison of outcomes data across international healthcare centers relies on the accurate capture of predictor variables such as extent of morbidity. Because standard outcome sets could be used among health institutions with no or different electronic medical record and administrative data structures, a systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[14] of studies about the development or subsequent validation of self-reported comorbidity assessments.

Literature Search

An exhaustive search strategy was developed in Embase.com by a medical librarian experienced in systematic review searches.[15] To retrieve articles about the validation of questionnaires on comorbidity, the search strategy combined thesaurus terms (Emtree terms for Embase and MeSH terms for Medline) with terms in the title or abstract for 3 elements: comorbidity, questionnaires, and validation or reliability.

The search strategy for Embase was optimized to find all potentially relevant terms and then translated to Medline (Ovid), Cochrane CENTRAL, and Web of Science Core Collection.[16] Additional references were retrieved from Google Scholar (the first 100 references as sorted by relevance), literature lists of relevant reviews, and included references. Abstracts needed to be in English, but there were no restrictions on the language of the manuscript or country of publication in the search strategy. The databases were last searched on March 5, 2018. The full search strategies for all the databases are included in the online Appendix in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.02.006.

Study Selection

The inclusion criterion for studies was a primary focus on the development or subsequent validation (eg, reliability and prediction/association with outcomes) of an instrument for collecting information on the presence of morbidity directly from the patient.

The exclusion criteria were:

- lack of *any* methodological description of instrument development (validation and/or reliability of instrument)
- description of the use of a patient-reported comorbidity questionnaire for risk-adjustment purposes or for deriving health utilities
- focus in the patient-reported morbidity instrument on a subset of specific conditions (based on nosologic criteria), thereby making the instrument not generalizable to a larger patient population (eg, a list of mental health comorbidities for psychiatric patients)

The search results were deduplicated[17] and then imported into Covidence (www.covidence.org, Melbourne, Australia), a Cochrane technology web-based platform developed specifically to screen and track articles through the inclusion and exclusion criteria process of a systematic review. In Covidence, the titles and abstracts of each reference were independently screened for relevance by 2 reviewers. The screening phase was conducted in the following order:

1. The first screening was based on the title and abstract. In the event that the article's aim did not meet the inclusion criteria but nevertheless mentioned a patient-reported medical comorbidity questionnaire, the full text was reviewed to determine the instrument used and references were further screened for any potential missing articles on the comorbidity questionnaire.
2. The second screening was based on the full-text assessment of retained articles. Studies that still did not meet our inclusion criterion were subsequently removed.

If the authors had any disagreements on article eligibility during the first screening, the study in question would be screened in a full-text version. Consensus on the inclusivity of selected articles was ultimately reached by the authors.

Included Comorbidities

The number of comorbidities, as well as a list of included comorbid conditions, was evaluated for every survey instrument. Additional survey questions, such as those evaluating the condition severity (impact on daily activities/functional status) or medication use for a comorbidity, were also noted.

Reliability

Measures of reliability (eg, test–retest reliability; patient-report vs other data sources such as medical records, administrative data, or laboratory testing) at either item-level or overall instrument level were catalogued for all studies. Because of the anticipated heterogeneity in the reporting between the studies, both reliability-specific values (intraclass correlation coefficient [ICC] and Kappa [κ] values) and other measures of the morbidity instrument's performance (Spearman correlation coefficient, sensitivity/specificity, and positive and negative predictive values) were included. The kappa values were measured based on the presence of the condition in the self-reported instrument versus the medical record (or administrative record), which was considered the gold standard. The articles mention that reporting a condition by self-report that is absent in the medical record could also imply a deficiency with the medical record. Kappa

values >0.80 indicate excellent agreement, 0.61 to 0.80 good agreement, 0.41 to 0.60 moderate agreement, 0.21 to 0.40 fair agreement, and < 0.20 poor agreement.[18] Spearman correlation coefficients are categorized as ≤0.20 (poor), 0.30 to 0.59 (fair), 0.60 to 0.79 (moderate), and > 0.80 (strong).[19]

Evaluated Outcomes

All instances where patient-reported morbidity instruments were used to assess association with or predict certain outcomes as part of the validation study were documented for this review. Examples of the outcome metrics included are mortality, disease response, patient-reported outcome measures (PROMs), adverse events and other events of interest, and healthcare utilization/costs, as per the categories of outcomes used in the Agency for Healthcare Research and Quality (AHRQ) Outcome Measures Framework (OMF).[20]

Questionnaire Length, Duration, Responsiveness, and Utilization

The length (number of items/questions) of the instrument and duration of completion was documented, when available, as was the route of administration (self-administration vs administration by a clinical or research associate). Finally, the number of times the paper had been cited in Web of Science was also noted.

Results

Included Studies

Figure 1 details the search and inclusion strategy. In total, 1005 studies met our search criteria; 70 studies met our criteria for inclusion in the full-text assessment. Thirty-six studies were eliminated after the full-text review, leaving 34 studies for inclusion in this systematic review.

A summarized overview of all the included articles in this systematic review is included in Table 1. Descriptive characteristics, reliability, validity, and evaluated outcomes of morbidity instruments are shown in Table 2.[21-67]

Included Patient-Reported Morbidity Instruments

Ten original patient-reported morbidity instruments were identified, with most of these development studies being conducted in the United States. The instruments considered originals were: the Self-Reported Charlson Comorbidity Index (SR-CCI),[11] the Self-Administered Comorbidity Questionnaire (SCQ),[27] the Disease Burden Morbidity Assessment (DBMA),[32] the Comorbidity Symptom Scale (CmSS),[38] the Patient Self-Administered Health History Questionnaire,[47] the Multi-Morbidity Assessment Questionnaire for Primary Care (MAQ-PC),[59] the Patient-Based Comorbidity Index (CI),[64] the Health Impact Index (HII),[50] the Seattle Index of Comorbidity (SIC),[48] and an unnamed prognostic index (including comorbidities).[60] The SR-CCI and SCQ instruments were the most frequently cited. Other included articles were translation and cross-cultural adaptation studies, variations of these questionnaires (eg, with a small number of items added or removed), or validation studies.

Presence of Specific Conditions and Related Assessments

Conditions that were most commonly included in the patient-reported morbidity instruments were diabetes, hypertension, myocardial infarction, and stroke (see Table 2). The question regarding the presence or absence of specific comorbidities was presented in multiple ways, including:

- "Do you have or have you ever had…?"[32,34,41,50,60,65]
- "Has a doctor ever told you that you have…?"[5,44,45,48,55,59,60]
- "Do have you any of the following problems?"[27,28,30,45]

Most instruments had close-ended response alternatives for each condition listed. Some questionnaires had an additional free-text item for patients to report additional comorbidities that were not listed in the instrument.[27,33,47,52,63]

Several instruments included additional questions regarding the severity of the conditions, such as, "Does it limit your activities?"[5,27,28,30,32,34] or regarding active treatment, for example, "Do you receive treatment for it?"[5,27,28,30,44]

Figure 1. Flowchart of relevant article selection.

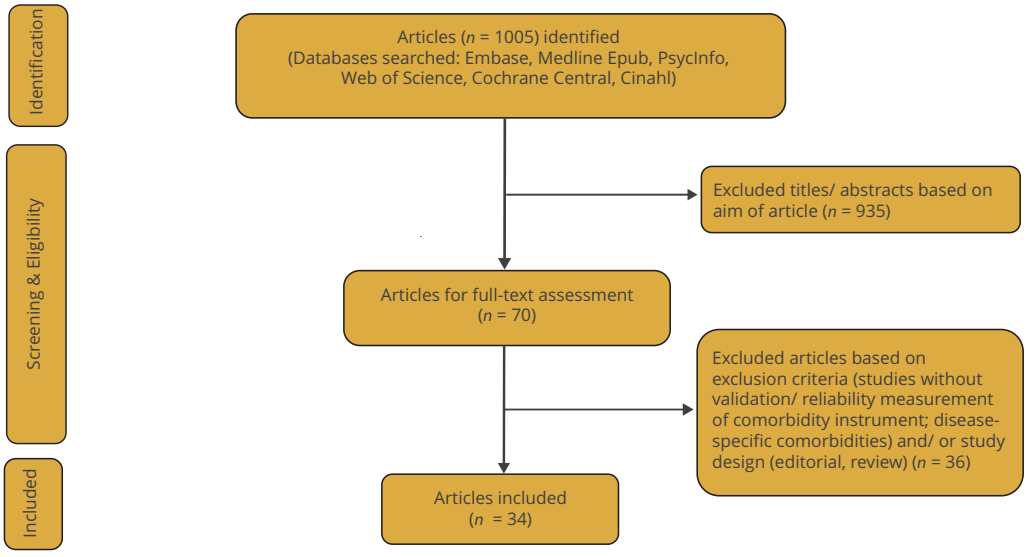

Table 1. Summarized overview of included studies.

Patient-reported morbidity instruments	Selected article and country of origin	Availability of reliability data	Evaluated outcomes	Method of questionnaire administration	Time to questionnaire completion	Number of Citations
Patient-Reported Charlson Comorbidity Index	Katz, 1996*,† United States	Item-level: + Overall: +	Mortality: − PROM: − Healthcare utilization: +	Self- or interviewer-administered	10 minutes	758
	Susser, 2008* Canada	Item-level: + Overall: +	Mortality: − PROM: + Healthcare utilization: −	Self-administered or filled out by proxy	-	19
	Corser, 2008* United States	Item-level: + Overall: +	Mortality: − PROM: − Healthcare utilization: −	Interviewer-administered	-	53
	Olomu, 2012* United States	Item-level: − Overall: −	Mortality: − PROM: + Healthcare utilization: −	Interviewer-administered	-	23
	Ng, 2015* Singapore	Item-level: + Overall: +	Mortality: − PROM: + Healthcare utilization: −	Self- or interviewer-administered	15 minutes	5
	Habbous, 2013* Canada	Item-level: + Overall: +	Mortality: + PROM: − Healthcare utilization: -	Self-administered	-	11
	Chaudhry, 2005* United States	Item-level: + Overall: −	Mortality: + PROM: − Healthcare utilization: +	Self-administered	1 minute	172
Self-Reported Comorbidity Questionnaire	Sangha, 2003*,† United States	Item-level: + Overall: +	Mortality: − PROM: + Healthcare utilization: +	Self-administered	-	757
	Stolwijk, 2014* Netherlands/ Belgium	Item-level: + Overall: +	Mortality: − PROM: + Healthcare utilization: −	Self-administered	-	17
	Robinski, 2016* Germany	Item-level: + Overall: −	Mortality: − PROM: + Healthcare utilization: −	Self-administered	-	2
Disease Burden Morbidity Assessment	Bayliss, 2005† United States	Item-level: + Overall: −	Mortality: − PROM: + Healthcare utilization: −	Self-administered	-	133
	Poitras, 2012 Canada	Item-level: + Overall: −	Mortality: − PROM: − Healthcare utilization: −	Self-administered	<15 minutes	19
	Wijers, 2017 Spain	Item-level: − Overall: −	Mortality: − PROM: + Healthcare utilization: −	Self-administered	-	2
	Simpson, 2004 United States	Item-level: + Overall: −	Mortality: − PROM: − Healthcare utilization: −	Self-administered	-	214
Comorbidity Symptom Scale	Crabtree, 2000* England	Item-level: − Overall: −	Mortality: − PROM: + Healthcare utilization: −	Interviewer-administered	<10 minutes	29

continued on next page

Table 1. Continued

Patient-reported morbidity instruments	Selected article and country of origin	Availability of reliability data	Evaluated outcomes	Method of questionnaire administration	Time to questionnaire completion	Number of Citations
	De-Loyde, 2015 Australia	Item-level: – Overall: –	Mortality: – PROM: – Healthcare utilization: –	Self-administered	-	10
	Gad, 2012 United States	Item-level: + Overall: –	Mortality: – PROM: – Healthcare utilization: –	Self-administered	-	3
	Hansen, 2014 Germany	Item-level: + Overall: –	Mortality: – PROM: – Healthcare utilization: –	Interviewer-administered	-	-
	Horton, 2010 Canada/United States	Item-level: + Overall: –	Mortality: – PROM: – Healthcare utilization: –	Self-administered	11 minutes (mean)	63
Questionnaire from CALAS study	Iecovich, 2013 Israel	Item-level: + Overall: +	Mortality: – PROM: – Healthcare utilization: –	Interviewer-administered	30 – 45 minutes	1
	Klabunde, 2006 United States	Item-level: – Overall: –	Mortality: – PROM: – Healthcare utilization: –	Self-administered	-	114
	Boissonnault, 2005 United States	Item-level: + Overall: +	Mortality: – PROM: – Healthcare utilization: –	Self-administered	-	15
Seattle Index of Comorbidity	Fan, 2002* United States	Item-level: – Overall: –	Mortality: + PROM: + Healthcare utilization: +	Self-administered	-	123
Health Impact Index	Lorem, 2016* Norway	Item-level: – Overall: +	Mortality: – PROM: + Healthcare utilization: –	Self-administered	-	4
	Lucke, 2016 Germany	Item-level: + Overall: +	Mortality: – PROM: – Healthcare utilization: –	Self-administered	-	12
Cornell Medical Index	Md Yusof, 2010* United Kingdom	Item-level: – Overall: +	Mortality: + PROM: – Healthcare utilization: –	Self-administered	-	2
Questionnaire from CHOICE study	Merkin, 2007 United States	Item-level: + Overall: –	Mortality: – PROM: – Healthcare utilization: –	-	-	70
Questionnaire from AHEAD study	Mukerji, 2007 United States	Item-level: + Overall: +	Mortality: – PROM: – Healthcare utilization: –	Self-administered or interviewer administered	-	41
	Paleri, 2002* United Kingdom	Item-level: + Overall: –	Mortality: – PROM: – Healthcare utilization: –	Self-administered	8.3 minutes	22
Multi-Morbidity Assessment Questionnaire for Primary Care	Pati, 2016 India	Item-level: + Overall: –	Mortality: PROM: Healthcare utilization:	-	20 – 25 minutes	5

continued on next page

Table 1. Continued

Patient-reported morbidity instruments	Selected article and country of origin	Availability of reliability data	Evaluated outcomes	Method of questionnaire administration	Time to questionnaire completion	Number of Citations
	Lee, 2006* United States	Item-level: − Overall: −	Mortality: + PROM: − Healthcare utilization: −	Interviewer-administered	-	416
Self-Report Comorbidity	Voaklander, 2004* Canada	Item-level: − Overall: +	Mortality: − PROM: + Healthcare utilization: +	-	-	12
(Patient-Based) Comorbidity Index	Selim, 2004* United States	Item-level: − Overall: −	Mortality: + PROM: + Healthcare utilization: +	Interviewer-administered	-	150
Questionnaire from LACE study	Vigen, 2016 United States	Item-level: + Overall: −	Mortality: + PROM: − Healthcare utilization: −	Self- or interviewer-administered	-	7

Note. + = described in study; − = unknown/not described in study
AHEAD indicates Action for Health in Diabetes; CALAS, Cross-Sectional and Longitudinal Aging Study; CHOICE, Choices for Healthy Outcomes in Caring for End-stage renal disease study; LACE, Life After Cancer Epidemiology study; PROM, patient-reported outcome measures.
*Instrument associated with an index
†Main developmental study

Instrument Administration, Length, Duration, and Responsiveness

Nineteen studies had self-administered questionnaires (either in mailed/written or electronic form),[5,25-28,30,32,33,34,36,39,41,44,47,48,50,52,54,57] whereas 8 studies had questionnaires that were administered verbally by a clinical/research associate (either face-to-face or by phone)[12,22,38,42,45,60,64] and 4 studies reported both administration methods.[9,11,65,24] Of the associate-administered surveys, it was not clear whether clarifying questions were allowed or used in nearly all of the studies. Five studies[9,11,24,57,65] had comorbidity questionnaires that could either be self-administered or administered by an interviewer (eg, associate or other proxy) if needed (eg, for patient illiteracy).

The length of the instruments varied from 4 items[65] to 195 items[54] (divided over multiple physical and mental sections). Nine studies mentioned the duration to complete the questionnaire, which ranged from 1 minute[26] to 45 minutes.[45] Response rates were provided in 9 studies ranging from 28%[32] to 99%.[30]

Reliability and Concordance with Other Data Sources

Test–retest reliability was described in 7 studies[5,11,27,33,38,47,59] (data not shown), mostly measured by the intraclass correlation or Spearman correlation coefficients. The amount of patients on which it was tested ranged from 26[11,27] to 103.[2] The interval period between both measurements varied from 24 hours[11,27,47] to 4 weeks.[38] The overall Spearman correlation coefficients for patient-reported comorbidity questionnaires ranged from 0.73[11] (moderate reliability) to 0.87 (strong reliability),[38] whereas the intraclass correlation coefficients ranged from 0.86[33] to 0.97.[59]

Whole instrument and item-level concordance of patient-reported morbidity scores with information from other data sources, either medical records or medical record-derived comorbidity indices, were most frequently assessed (see Table 2). Spearman correlation coefficients for the relationship between patient-reported morbidity scores and composite scores from other data sources ranged from r = 0.24 (14 conditions)[28] to r = 0.70 (18 conditions).[11] In studies measuring Kappa coefficients, κ values were notably higher for agreement with medical records (κ range: 0.56-0.69)[44,47] as opposed to agreement with medical record-derived morbidity indices (κ range: 0.37-0.50).[25,57] Administrative data were also used as a comparative data source in a number of studies,[9,26,45,52,63] which generally demonstrated poor agreement.[9,45,63]

Item-level (single condition) was the most commonly reported form of concordance assessment, mostly measured against medical records or derived morbidity indices. A striking observation was that diabetes, as a comorbidity questionnaire item, had the highest Kappa value across included studies.[24,26,30,39,41,42,44,45,55,57,65] Most included studies had a substantially wide κ value range,[22,24,25,26,27,28,30,36,41,42,44,45,55,57] in general from 0.66[27] to 0.86.[28]

Association with Health and Healthcare Outcomes

A number of studies assessed the association between patient-reported instrument scores and mortality, patient-reported outcome measures, and healthcare utilization. Because none of the included studies evaluated their instrument against all 3 outcomes, we provided some examples in this paragraph to demonstrate the directionality of the associations.

Some studies assessed the relationship between patient-reported instrument scores and mortality or survival.[25,26,48,54,60,64,65] Habbous et al[25] demonstrated a significant relation between the patient-reported Charlson Comorbidity Index (CCI) and overall survival, with the presence of at least 2 comorbidities being associated with worse survival (hazard ratio [HR] = 1.62, *P*=.003). Nevertheless, this relation was stronger for the nonpatient-reported, medical record-derived CCI (HR = 2.60). Only a few studies developed prediction models for all-cause mortality with patient-reported comorbidity instruments, either in combination with other predictors such as demographic variables[60] or by themselves.[26,48] Fan et al[48] developed a prediction model with the Seattle Index of Comorbidity (SIC) and estimated an area under the curve (AUC) = 0.71 for all-cause mortality at 2 years follow-up, whereas Lee et al[60] estimated an AUC = 0.82 of a different model (including sex,

Table 2. Detailed overview of included studies.

	Items	Study population (n)	Item-level reliability vs other data sources	Overall instrument reliability vs other data sources	Evaluated outcome(s)
Katz[11]	AIDS Any tumor Cerebrovascular disease Chronic pulmonary disease Congestive heart failure Connective tissue disease Dementia Diabetes (end-organ damage) Diabetes (mild to moderate) Hemiplegia Leukemia Liver disease Lymphoma Metastatic tumor Moderate/severe renal disease Myocardial infarction Peripheral vascular disease Ulcer disease	170 inpatients from 6 care units (3 medical and 3 surgical) at 1 hospital. Characteristics: Female 55% Mean age 65.3 years (± SD 8.8) Caucasian 82% College level or higher 50% Surgical 54%	Comparison: Medical record-derived CCI Results: Kappa: 0.35 (ulcer disease/ diabetes with end-organ damage) – 0.85 (leukemia) Sensitivity: Specificity: - PPV: - NPV: - Agreement between self-reported CCI and medical record-derived CCI ranged from 83% (any tumor) to 100% (AIDS).	Comparison: Medical record-derived CCI Results: Self-reported CCI score was higher versus medical-record derived CCI (1.99 ± SD 2.13 vs 1.59 ± SD 1.80, $P < .01$) Spearman r range 0.63 ($P = .0001$) (full index) - 0.70 ($P = .0001$) (when the solid tumor item was excluded from the analysis)	Measured in 49 and 56 patients on medical and surgical service respectively. Results: Mortality: - PROM: - Healthcare utilization: Hospitalizations in last year: Spearman r range 0.17-0.31, $P < .05$ Number of prescription medication: Spearman r range 0.26-0.44, $P < .05$ Hospital charges during admission: Spearman r range 0.09-0.26, $P < .05$) Length of stay: Spearman r range 0.15-0.20
Susser[9]	As per Katz	520 elderly patients ready to be discharged home from the ER. Data from a previously published RCT.[21] Characteristics: Female 60% Age group >75 years 57%	Comparison: Administrative data-derived CCI Results: Kappa: Highest $\kappa = 0.55$ (chronic pulmonary disease). Individual Kappa values with range were not described for all conditions. Sensitivity: - Specificity: - PPV: - NPV: - Four conditions were reported more frequently by self-report (myocardial infarction, ulcer disease, diabetes with end-organ damage, and connective tissue disease), whereas 5 (hemiplegia, mild-moderate diabetes, solid tumor, lymphoma, and dementia) were more frequently reported in administrative data.	Comparison: Administrative data-derived CCI Results: Poor agreement between self-reported and administrative data-derived CCI, indicated by an (overall) ICC = 0.43 (95% CI 0.40-0.47).	Comparison: Administrative data-derived CCI Results: Mortality: - PROM: ADL (functional decline): predictive ability of self-reported vs administrative data-derived CCI was measured with unweighted (for sampling) AUC = 0.51 vs AUC = 0.54 and weighted AUC = 0.54 vs AUC = 0.50, $P > .05$) Healthcare utilization: Hospital days: self-reported vs administrative data-derived CCI was measured with unweighted AUC = 0.63 vs AUC = 0.63 and weighted AUC = 0.68 vs AUC = 0.69, $P > .05$) ER visits: unweighted AUC = 0.64 vs AUC = 0.65 and weighted AUC = 0.67 vs AUC = 0.63, $P > .05$)
Corser[22]	As per Katz	525 patients admitted for acute coronary syndrome in 5 hospitals. Characteristics: Female 36.4% Mean age 59.73 years (± SD 12) Caucasian 84.4% College level or higher 43.8%	Comparison: Medical record-derived CCI Results: Kappa: 0.07 (CTD, RA) –0.80 (diabetes). Only conditions with a prevalence of at least 3% (in each data source) were included in the Kappa analysis. Sensitivity: - Specificity: - PPV: - NPV: -	Comparison: Medical record-derived CCI Results: Self-reported CCI (composite) scores were higher than medical record-derived CCI scores (mean 1.78 ± SD 1.99 vs mean 1.27 ± SD 1.43). Correlation between self-reported and medical record-derived CCI composite scores were fair (Spearman r = 0.57, $P < .01$).	-

continued on next page

Table 2. Continued

	Items	Study population (n)	Item-level reliability vs other data sources	Overall instrument reliability vs other data sources	Evaluated outcome(s)
Olomu[12]	As per Katz	525 patients admitted for acute coronary syndrome in 5 hospitals. Data from a previously published RCT.[23] Characteristics: Female 36.4% Caucasian 84.4% College level or higher 43.8%	-	-	Comparison: Medical record-derived CCI. Results: Mortality: - PROM: ASI (functional capacity): Prediction at 3 months was slightly better with SCQ vs the CCI (R^2 = 0.340, $P <$ .0005 vs R^2=0.331, $P <$ .0035), whereas it was slightly better with CCI vs SCQ (R^2 = 0.370, $P <$ .0005 vs R^2=0.358, $P <$ .0005) at 8 months. EQ-5D (health-related quality of life): Only the SCQ significantly predicted EQ-5D scores at 3 and 8 months (R^2 = 0.288 and R^2 = 0.265, $P <$ .0005), whereas the CCI did not (R^2 = 0.262, $P >$ .201 and R^2= 0.245, $P >$ .132). Healthcare utilization: -
Ng[24]	As per Katz	301 rheumatic patients from 1 tertiary hospital. Characteristics: Female 61.5% Median age 51 years (21-79) Chinese 68.8% College level or higher 54.7%	Comparison: Medical record-derived CCI Results: Kappa: 0.189 (diabetes with end-organ damage) – 0.764 (diabetes). Kappa values was only calculated for 8 of 18 conditions that did not have any cell values of zero. Sensitivity: 33.3 (diabetes with end-organ damage) – 100% (myocardial infarction) Specificity: 58.9 (CTD, RA) – 99.1% (CVA) PPV: - NPV: - Agreement between self-reported CCI and medical record-derived CCI ranged from 74.1% (CTD/ RA) to 100% (leukemia, lymphoma, metastatic solid tumor, AIDS).	Comparison: Medical record-derived CCI Results: Median self-reported composite CCI scores were higher than the medical record-derived CCI scores, indicating that conditions were generally reported more frequently by self-report than EHR review. Self-reported composite CCI scores had moderate agreement (ICC = 0.513, $P <$ .001) and strong correlation (Spearman r = 0.570, $P <$ .001) with the medical record-derived CCI scores.	Comparison: Medical record-derived CCI Results: Mortality: - PROM: SF-36 (health-related quality of life): Self-reported CCI was negatively associated with PCS (β = –2.56, $P <$ .001) and MCS (β = –1.24, P = .044). Medical record-derived CCI scores had a similar trend but coefficients didn't reach statistical significance. Healthcare utilization: -
Habbous[25]	*Exposures* Smoking and alcohol *Conditions/Diseases* Chronic cough/bronchitis Dementia (eg, Alzheimer's) Diabetes (eye/kidney problems) (Past) Dialysis requirement Emphysema Heart failure Hemiplegia Hepatitis HIV/AIDS Liver disease Myocardial infarction Other joint/bone problems Past cancer history Peripheral vascular disease Rheumatoid arthritis Serious kidney problems Stomach ulcers (test-proven) Stroke/mini-stroke	882 head-and-neck cancer patients. Characteristics: Female 23% Median age 61.5 years (61-62.5) Caucasian 84%	Comparison: Medical record-derived CCI Results: Kappa: 0.16 (hemiplegia) – 0.93 (diabetes) Sensitivity: - Specificity: - PPV: - NPV: - Positive agreement between self-reported CCI and medical record-derived CCI ranged from 17% (hemiplegia) to 94% (diabetes). Negative agreement between self-reported CCI and medical record-derived CCI ranged from 84 (CTD) to 100% (dementia).	Comparison: Medical record-derived CCI Results: Patient-reported CCI scores were higher than the medical record-derived CCI scores (mean 1.01 (95% CI 0.9-1.1) vs 0.74 (95% CI 0.7-0.8), $P <$ .0001). Comorbidities were reported more often by patients in comparison to medical records review. Overall agreement between patient-reported CCI and medical record-derived CCI was measured as κ = 0.37, which improved if CTD (κ = 0.52) or COPD (κ = 0.43) was removed from the patient-reported CCI score.	Comparison: Medical record-derived CCI Results: Mortality: Overall survival: Both patient-reported CCI (HR 1.62 (95% CI 1.18-2.24), P = .003) and medical record-derived CCI (HR 1.97 (95% CI 1.38-2.80), P = .0002) were significantly associated with overall survival after multivariate (age, sex, marital status, stage, and disease site) adjustment, when at least 2 comorbidities were present. PROM: - Healthcare utilization: -

continued on next page

Table 2. Continued

	Items	Study population (n)	Item-level reliability vs other data sources	Overall instrument reliability vs other data sources	Evaluated outcome(s)
Chaudhry[26]	Asthma/emphysema/ Chronic bronchitis Arthritis or rheumatism Cancer (diagnosed within past 3 years) Diabetes Digestive problems (ie, ulcer/ colitis/gallbladder disease) Heart trouble (ie, angina/ CHF/ CAD) HIV or AIDS Kidney disease Liver problems (cirrhosis) Stroke	7761 hospitalized general medicine patients at a single center. Characteristics: Female 62% Mean age 56-57 years African-American >80% MMSE score > 17	Comparison: Administrative data-derived CCI Results: Item-level data vs one-year look-back: Kappa: 0.04 (stomach ulcer) – 0.83 (diabetes) Sensitivity: 44 (cancer) – 86% (diabetes, HIV/ AIDS) Specificity: 48 (arthritis or rheumatism) – 98% (HIV/ AIDS) PPV: 3 (stomach ulcers) – 90% (diabetes) NPV: 91 (asthma, emphysema, or bronchitis) – 100% (HIV/ AIDS) Item-level data vs index hospitalization: Kappa: 0.06 (arthritis or rheumatism) – 0.82 (diabetes) Sensitivity: 43 (cancer) – 91% (diabetes) Specificity: 47 (arthritis or rheumatism) – 95% (liver disease, cancer) PPV: 9 (arthritis or rheumatism) – 84% (diabetes) NPV: 93 (asthma, emphysema or bronchitis) – 99% (heart disease, kidney disease, liver disease, cancer) No statistically significant differences in Kappa values, sensitivities, specificities, or positive or negative predictive values was observed for 1-year look-back periods or index hospitalization.	-	The predictive power of the *self-reported* CCI was constructed with 4 different logistic regression models performed in a validation cohort (n = 3870). Model 1: age, sex + original CCI weight Model 2: age, sex + study-specific CCI weight Model 3: age, sex, diagnosis-related group weight + original CCI weight, Model 4: age, sex, diagnosis-related group weight + study-specific CCI weight Results: Mortality: One-year mortality: AUCs for the self-reported CCI were 0.70 (0.68-0.73) for model 1, 0.72 (0.70-0.75) for model 2, 0.75 (0.72-0.77) for model 3, and 0.76 (0.73-0.78) for model 4. AUCs were slightly less compared with the administrative data-derived CCI indices ($P < .001$). PROM: - Healthcare utilization: Log total costs: R^2 values for the different regression models ranged from 0.02 (models 1 & 2) – 0.33 (models 3 & 4) Log length of stay: R^2 values for the different regression models ranged from 0.01 (model 1) – 0.22 (models 3 & 4)

continued on next page

Table 2. Continued

	Items	Study population (n)	Item-level reliability vs other data sources	Overall instrument reliability vs other data sources	Evaluated outcome(s)
Sangha[27]	Anemia or other blood disease Back pain Cancer Depression Diabetes Heart disease High blood pressure Kidney disease Liver disease Lung disease Osteoarthritis/ Degenerative arthritis Other medical problems (optional) Rheumatoid arthritis Ulcer or stomach disease	170 hospitalized patients from 6 care units at one hospital. Characteristics: Female 55% Mean age 65.3 years (± SD 8.8) Caucasian 82% College level or higher 50%	Comparison: Medical record-derived CCI Results: Kappa: 0.27 (lung disease) – 0.93 (liver disease) Sensitivity: - Specificity: - PPV: - NPV: - Overall agreement between the SCQ and the medical record-derived CCI ranged from 78% (heart disease) to 99% (liver disease).	Comparison: Medical record-derived CCI Results: SCQ scores were higher than the medical record-derived CCI scores (mean 5.61 ± SD 4.1 vs 1.59 ± SD 2.13) SCQ had a fair correlation (Spearman r = 0.32) with the medical record-derived CCI, which slightly increased (Spearman r = 0.55) when questionnaires were truncated to only comparable items.	Comparison: Medical record-derived CCI Results for medical patients: Mortality: - PROM: - SF-36 (health-related quality of life): SCQ had poor to modest correlations with SF-36 (subscale) scores at one-year follow-up, ranging from "MCS" (Spearman r = –0.03, $P > .05$) to "General health" (Spearman r = –0.39, $P < .0001$). Total SCQ scores explained substantial variation for most SF-36 subscales, with R^2 values ranging from 0.10 ("Social function") to 0.25 ("Physical function") in multivariate (including age, sex, ethnicity, education level, and insurance status) linear regression models. Healthcare utilization: Hospitalizations in previous year: SCQ scores correlated fairly with hospitalizations in the previous 12 months (Spearman r = 0.21, $P < .01$ for medical patients; Spearman r = 0.37, $P < .01$ for surgical patients) Prescription medication: SCQ scores also correlated moderately with number of prescriptions (Spearman r = 0.40 for medical patients; r = 0.55 for surgical patients) Total hospital charges: SCQ scores correlated poorly with total inpatient charges (Spearman r = 0.09 for medical patients; r = 0.10 for surgical patients) Length of stay: SCQ scores also correlated poorly with hospital length of stay (Spearman r = 0.03 for medical patients; r = 0.14 for surgical patients)

continued on next page

Table 2. Continued

	Items	Study population (n)	Item-level reliability vs other data sources	Overall instrument reliability vs other data sources	Evaluated outcome(s)
Stolwijk[28]	Dutch modified version (mSCQ) of the SCQ instrument[27] Anemia or other blood disease Back pain Cancer Depression Diabetes Heart disease High blood pressure Kidney disease Liver disease Lung disease Osteoarthritis Other non-specified medical problems (optional; max. 3) Rheumatoid arthritis Ulcer or stomach disease	98 outpatients with ankylosing spondylitis. Data from the OASIS study.[29] Characteristics: Female 29.6% Mean age 53.9 years (± SD 11.4) College level or higher 15.7%	Comparison: Medical records Results: Kappa: 0.14 (osteoarthritis, ulcer disease) – 1.00 (cancer). Kappa analysis included 10 conditions. Sensitivity: - Specificity: - PPV: - NPV: -	Comparisons: 1. Medical record-derived CCI 2. Michaud-Wolfe index Results: SCQ had poor to fair correlations with the medical record-derived CCI (Spearman r = 0.24, $P < .05$) and Michaud-Wolfe index (Spearman r = 0.43, $P < .05$) mSCQ also had moderate correlations with CCI (Spearman r = 0.36, $P < .05$) and Michaud-Wolfe index (Spearman r = 0.57, $P < .05$)	Comparisons: 1. Medical record-derived CCI 2. Michaud-Wolfe index Results: Mortality: - PROM: BASDAI (disease activity): SCQ correlated moderately with disease activity (Spearman r = 0.27, $P < .05$), while CCI correlated poorly (Spearman r = 0.01). SCQ was significantly associated (OR = 1.73, 95% CI 1.25-2.40, $P < .01$) with low disease activity (BASDAI < 4). BASFI (physical function): SCQ correlated moderately (Spearman r = 0.43, $P < .05$) with physical function, but was significantly associated (β = 0.11, 95% CI 0.03-0.19, $P = .01$) with BASFI. SF-36 (health-related quality of life): SCQ correlated moderately (Spearman r = -0.45, $P < .05$) with the PCS subscale, and was significantly associated (β = –0.72, $P = .03$) with PCS. CCI and Michaud-Wolfe indices were not significantly associated with BASFI and SF-36-PCS. Healthcare utilization: -
Robinski[30]	German version (SCQ-G) of the SCQ instrument[27]	780 adult end-stage renal disease patients from 55 dialysis units. Data from the CORETH project.[31] Characteristics: Female 32.6% Mean age 63.2 years (± SD 15.1)	Comparison: Medical record-derived CCI Results: Kappa: 0.01 (peptic ulcer disease) – 0.84 (diabetes) Sensitivity: - Specificity: - PPV: - NPV: - Overall agreement between the SCQ and CCI ranged from 70% (heart disease) to 95% (kidney disease, liver disease). Positive agreement between both data sources ranged from 6% (peptic ulcer disease) to 97% (kidney disease). Negative agreement ranged from 78% (heart disease) to 97% (liver disease).	Comparison: Medical record-derived CCI Results: Total SCQ-G score was moderately correlated to the medical record-derived CCI (Spearman r = 0.27, $P < .01$).	Comparison: Medical record-derived CCI Results: - Mortality: - PROM: - SF-12 (health-related quality of life): Total SCQ-G score was moderately correlated with MSC (Spearman r = –0.25, $P < .01$) and PSC (Spearman r = –0.49, $P < .01$) subscales, while CCI correlated poorly with MSC (Spearman r = 0.06, $P > .05$) and moderately with PSC (Spearman r = –0.36, $P < .01$). Healthcare utilization: -

continued on next page

Table 2. Continued

	Items	Study population (n)	Item-level reliability vs other data sources	Overall instrument reliability vs other data sources	Evaluated outcome(s)
Bayliss[32]	Angina/ CAD Asthma Back pain (chronic) or sciatica Bronchitis (chronic)/ COPD Cancer (diagnosed within past 5 years) Cholesterol (elevated) Colon problem (eg, diverticulitis/ irritable bowel) Congestive heart failure Diabetes Hard of hearing Hypertension Kidney disease Nerve condition Osteoarthritis Osteoporosis Overweight Poor circulation (eg, peripheral vascular disease) Rheumatic disease (eg, fibromyalgia or lupus) Rheumatoid arthritis Stomach problem (eg, gastritis/ ulcer/ reflux) Stroke Thyroid disorder Vision problem	156 patients (≥ 65 years) from the HMO Characteristics: Female 49.4% Mean age 75 years (67-94) Caucasian 91% College level or higher 59.6%	Comparison: Medical records Results: Kappa: - Sensitivity: 35 (kidney disease) – 100% (asthma) Specificity: 61 (hard of hearing) – 100% (kidney disease, cancer) PPV: - NPV: -	Comparison: Medical records Results: Sensitivity by respondent analysis: 14%-100% (median 75%) Specificity by respondent analysis: 59%-100% (median 91%)	Comparisons: 1. Medical records 2. Medical record-derived CCI 3. Rx-risk score (comorbidity measure including age, gender, health insurance benefit status, and a category based on diagnoses from administrative pharmacy data) Results: Mortality: - PROM: SF-36 (health-related quality of life): Self-reported number of conditions (Spearman r = 0.477, $P < .001$) had a similar correlation compared to the medical-record CCI (r = 0.48, $P < .001$) but higher compared to the Rx risk score (0.17, $P = .037$). SF-36 (physical functioning): Self-reported conditions (r = –0.482, $P < .001$) had a stronger correlation vs CCI (r = –0.41, $P < .001$) and the Rx Risk score (r = –0.18, $P = .035$). BRFSS (depression screening): Self-reported conditions (r = –0.24, $P = .003$) had a stronger correlation compared with CCI (r = –0.12, $P = .14$) and the Rx Risk score (r = –0.05, $P = .559$). GSE (self-efficacy): Self-reported conditions (r = –0.305, $P < .001$) had a stronger correlation vs CCI (r = –0.14, $P = .096$) and the Rx Risk score (r = 0.10, $P = .234$). Healthcare utilization: -
Poitras[33]	French modified version (DBMA-Fv) of the DBMA instrument[32]: 21 of the 23 original conditions were chosen. Items "Kidney disease" and "Nerve condition" were excluded in this version Item "Depression/ anxiety" was added to this version	78 patients from 1 health center. Characteristics: Female 68% Mean age 47.4 years (± SD 15.9) College level or higher 70.5%	Comparison: Medical records Results: Kappa: - Sensitivity: 62.5 (angina/ CAD) – 90% (diabetes) Specificity: 77.6 (overweight) – 98.6% (diabetes) PPV: 44.4 (overweight) – 92.9% (hypercholesterolemia) NPV: 88.7 (osteoarthritis) – 95.9% (asthma/ diabetes)	Comparison: Cumulative Illness Rating Scale Results: DBMA-Fv correlated moderately with the CIRS at baseline (r = 0.46, 95% CI 0.26-0.62, $P < .01$) and at 2 weeks' follow-up (Spearman r = 0.56, 95% CI 0.38-0.70, $P < .01$). Comparison: Medical records Results: Mean sensitivity of patient-reported conditions vs medical record review at 2 weeks was 73.9% (± SD 8.4), whereas mean specificity was 92.2% (± SD 6.7).	-

continued on next page

Table 2. Continued

	Items	Study population (n)	Item-level reliability vs other data sources	Overall instrument reliability vs other data sources	Evaluated outcome(s)
Wijers[34]	Spanish modified version of the DBMA instrument[32]: 21 out of the 23 original conditions were chosen Item "Liver disease" was excluded from this version due to low prevalence Items "UTI," "anxiety" and "memory-related disorders" were added to this version due to high prevalence in older adults	707 community-dwelling adults (≥65 years). Data from the ELES-PS study.[35] Characteristics: Female 57% Mean age 74.2 years (± 6.6) College level or higher 17.3%	-	-	Comparison: Self-reported conditions Results: - Mortality: - PROM: - PWI (physical functioning & perceived health status): DBMA significantly correlated stronger to physical functioning than self-reported number of conditions (Spearman r = –0.56 vs r = –0.51, *P* = .0035) PWI (quality of life): DBMA significantly correlated stronger to PWI in comparison to self-reported number of conditions (Spearman r = –0.41 vs r = –0.35, *P* = .0006) CES-D (depression screening): DBMA significantly correlated stronger to CES-D compared to self-reported number of conditions (Spearman r = 0.41 vs r = 0.35, *P* = .0043) Healthcare utilization: -
Simpson[36]	Angina pectoris Arthritis (OA/RA) Cancer Congestive heart failure Diabetes mellitus Disc disease Hip fracture Lung disease Myocardial infarction Osteoporosis Parkinson's disease Peripheral arterial disease Spinal stenosis Stroke	1002 disabled women, aged ≥65 years, with MMSE ≥18. Data from the Women's Health and Aging Study I.[37] Characteristics: Age group 65-74 years 44.2% Caucasian 71.1%	Comparison: Disease-specific standardized algorithms (medical history, physical examination, medication review, medical record review, radiographs, physician questionnaire) Results: Kappa: 0.24 (peripheral arterial disease) – 0.96 (hip fracture) Sensitivity: 22 (spinal stenosis) – 98% (stroke) Specificity: 45 (arthritis) – 100% (hip fracture, Parkinson's disease, disc disease, spinal stenosis) PPV: 0.20 (peripheral arterial disease) – 1.0 (Parkinson's disease) NPV: 0.38 (arthritis) – 1.0 (hip fracture, Parkinson's disease, cancer, stroke)	-	-

continued on next page

Table 2. Continued

	Items	Study population (n)	Item-level reliability vs other data sources	Overall instrument reliability vs other data sources	Evaluated outcome(s)
Crabtree[38]	Angina Anxiety and depression Any other condition Arthritis/osteoporosis Breathlessness secondary to cardiovascular cause Breathlessness/ Wheeze (secondary to respiratory cause) Cerebrovascular disease Constipation Cough/Sputum (secondary to COPD/ Asthma) Diabetes Diarrhea Epilepsy Hearing problems Pain Parkinson's disease Peripheral vascular disease Side effects from medications Skin disease Unsteadiness, falls, and syncope Upper gastrointestinal symptoms Urinary problems Visual problems Walking and mobility	183 patients ≥ 65 years with confirmed age-related cataract (n = 161) or from a geriatric day hospital (n = 22)	-	-	Comparison: - Results: Mortality: - PROM: - NEADL (activities of daily living): CmSS correlated moderately to the NEADL (Spearman r = 0.56, $P <$ .01). GHQ-28 (perceived health status): CmSS correlated poorly to the GHQ-28 (Spearman r = 0.48, $P <$ .01) HAD (anxiety and depression): CmSS correlated moderately to the HAD (Spearman r = 0.52, $P <$ .01). Healthcare utilization: -
De-loyde[39]	Another cancer Chronic respiratory disease Depression Diabetes Heart disease Hypertension Kidney disease	756 patients with colorectal cancer from multiple hospitals. Data from the CONNECT[40] RCT. Characteristics: Female 56% Age group <70 years 54% College level or higher 24%	Comparison: Clinician report Results: Kappa: 0.22 (another cancer) – 0.58 (diabetes) Sensitivity: - Specificity: - PPV: - NPV: - Comparison: Medical records Results: Kappa: 0.34 (kidney disease) – 0.77 (diabetes) Sensitivity: - Specificity: - PPV: - NPV: -	-	-
Gad[41]	Amputation Anemia or other blood problems Asthma/ Other lung disease Back pain Blood clots or phlebitis Bowel problems Cancer Chronic skin condition Congestive heart failure Depression or anxiety Diabetes Excessive weight Hearing loss Heart attack High blood pressure High cholesterol Kidney or urinary problems Liver/ Gallbladder disease Lupus/ Other autoimmune disease Neuromuscular disease Osteoarthritis/ Degenerative arthritis Osteoporosis Paralysis Peripheral vascular disease Previous fracture(s) Recent unwanted weight loss Rheumatoid arthritis Sleep problems Stroke Thyroid problems Ulcer/stomach problems Visual problems	382 preoperative orthopedic patients (aged ≥ 65 years) before undergoing total knee or hip arthroplasty Characteristics: Female 65% Mean age 74 years (± SD 6.1)	Comparison: Medical records Results: Kappa: 0.00 (osteoarthritis) – 0.76 (diabetes) Sensitivity: 9 (peripheral vascular disease) – 71% (hypertension) Specificity: 44 (osteoarthritis) – 99% (diabetes) PPV: - NPV: -	-	-

continued on next page

Table 2. Continued

	Items	Study population (n)	Item-level reliability vs other data sources	Overall instrument reliability vs other data sources	Evaluated outcome(s)
Hansen[42]	For analysis (32 of 46 diagnosis groups): Anemia Asthma/ COPD Atherosclerosis/ PAOD Cancers Cardiac arrhythmias Cardiac insufficiency Cardiac valve disorders Cerebral ischemia/ Chronic stroke Chronic cholecystitis/ Gallstones Chronic ischemic heart disease Chronic low back pain Diabetes mellitus Dizziness Gynecological problems Hemorrhoids Hypertension Hyperuricemia/gout Intestinal diverticulosis Joint arthrosis Lipid metabolism disorders Lower limb varicosis Migraine/chronic headache Neuropathies Osteoporosis Parkinson disease Prostatic hyperplasia Psoriasis Renal insufficiency Rheumatoid arthritis/ chronic polyarthritis Severe vision reduction Thyroid dysfunction Urinary tract calculi	3189 multimorbid primary care patients. Data from the Multi-Care Cohort Study.[43] Characteristics: Female 59.3% Mean age 74.4 years (± SD 5.2) College level or higher 10.9%	Comparison: Clinician report Results: Kappa: 0.05 (gynecological problems) – 0.80 (diabetes) Sensitivity: - Specificity: - PPV: - NPV: -	-	-
Horton[44]	Anemia Anxiety Arthritis Bipolar disorder Breast cancer Cataracts Colon cancer Depression Diabetes Epilepsy Fibromyalgia Glaucoma Heart disease Hip replacement Hyperlipidemia Hypertension Inflammatory bowel disease Irritable bowel syndrome Kidney disease Knee replacement Liver disease Lung cancer Lung disease Migraine Osteoporosis Peptic ulcer disease Peripheral vascular disease Rectal cancer Rheumatoid arthritis Schizophrenia Sjogren's syndrome Skin cancer Systemic lupus erythematosus Thyroid Uveitis Vitamin-B12 deficiency	404 patients with multiple sclerosis from 2 centers. Characteristics: Female 76% Mean age 46.5 years (± SD 11.8) Caucasian 92% College level or higher 63.4% Relapsing-remitting MS 70.8%	Comparison: Medical records Results: Kappa: 0.19 (anemia) – 0.88 (diabetes) Sensitivity: 14 (kidney disease) – 100% (bipolar disorder, breast cancer, glaucoma, lung cancer, rheumatoid arthritis, schizophrenia, cataracts) Specificity: 87 (depression) – 100% (breast cancer, lung cancer) PPV: 0.07 (skin cancer) – 1.00 (breast cancer/ lung cancer) NPV: 0.84 (depression) – 1.00 (bipolar disorder, breast cancer, cataracts, glaucoma, lung cancer, rheumatoid arthritis, schizophrenia)	Comparison: Medical records Results: Agreement between self-reports and medical records was κ = 0.56 (95% CI 0.48-0.64) for the presence of any physical comorbidity, and κ = 0.57 (95% CI 0.48-0.65) for mental comorbidities. For this analysis, the questionnaire was divided into physical vs mental comorbid conditions, and thereafter dichotomized in 0 vs >0 comorbidities.	-

continued on next page

Table 2. Continued

	Items	Study population (n)	Item-level reliability vs other data sources	Overall instrument reliability vs other data sources	Evaluated outcome(s)
Iecovich[45]	Arthritis Cancer Cardiovascular accident Circulatory disease Diabetes Gastrointestinal disease Hypertension Myocardial infarction Osteoporosis Other heart diseases Renal problems Respiratory disease Thyroid disease	402 disabled older patients who used adult daycare centers. Characteristics: Female 74.8% Mean age 78 years (± SD 7.02) Asian/ African 62.6% College level or higher 10.1%	Comparison: Medical records (including diagnostic ICD-9 codes) Results: Kappa: 0.09 (circulatory disease) – 0.76 (diabetes) Sensitivity: 22.5 (cancer) - 79.1% (diabetes) - Specificity: 73.5 (renal) – 98% (cancer) PPV: 0.36 (circulatory disease) – 0.92 (hypertension) NPV: 0.42 (hypertension) – 0.87 (thyroid disorder)	Comparison: Medical records (including diagnostic ICD-9 codes) Results: Self-reports correlated fairly with the EHR (r = 0.45, $P < .001$).	-
Klabunde[5]	Angina Arthritis or rheumatism Chronic Lung Disease/ Bronchitis/emphysema Cirrhosis/liver disease Congestive heart failure Depression or anxiety Diabetes Hypertension IBD/colitis/Crohn disease Myocardial infarction Stroke/brain hemorrhage Stomach ulcers with bleeding	3095 prostate cancer survivors. Data from the PCOS study.[46] Characteristics: Age group >65 years 64% Caucasian 78% College level or higher 60%	-	-	-
Boissonnault[47]	Anemia Ankylosing spondylitis Arterial blockage of legs Asthma Cancer Chemical dependency Deep venous thrombosis Degenerative osteoarthritis or wear-and-tear arthritis Depression Diabetes (diagnosed after age 18 years) Diabetes (diagnosed before age 18 years) Emphysema Endometriosis Epilepsy/seizures Gout Headaches (>1 per week) Heart attack Heart valve problems Hepatitis Hypertension Hyperthyroid Hypothyroid Infections Multiple sclerosis Osteoporosis Other illnesses (please list) Rheumatoid arthritis Stomach/duodenal ulcers Stroke Tuberculosis Urinary incontinence Questionnaire contains 91 items divided into 8 sections (comorbidities, surgeries, medication, substance use and demographic characteristics)	100 preoperative orthopedic surgery patients at 1 hospital. Characteristics: Female 54% Mean age 46.9 years (± SD 16.7) College level or higher 64%	Comparison: NP/PA responses to identical questionnaire after medical record review and/or patient interview Results: Kappa: 0.15 (other illnesses) – 1.00; mean $\kappa = 0.69$ Sensitivity: - Specificity: - PPV: - NPV: -	Comparison: NP/PA responses to identical questionnaire after medical record review and/or patient interview Results: Mean percentage agreement across all questionnaire items between self-report and NP/PA report was 96%.	-

continued on next page

Table 2. Continued

	Items	Study population (n)	Item-level reliability vs other data sources	Overall instrument reliability vs other data sources	Evaluated outcome(s)
Fan[48]	Angina Arthritis CABG/PTCA Cancer Congestive heart failure Coronary artery disease Depression Diabetes Drug abuse Enlarged prostate Heartburn HIV Hypertension Liver disease Lung disease Osteoporosis Pneumonia Post-Traumatic Stress Disorder Prior myocardial infarction Renal insufficiency Seizure Stroke Thyroid disease Ulcer disease	Development sample: 5469 patients from 7 VA medical centers. Data from the ACQUIP[49] study. Characteristics: Female 2.5% Mean age 67.8 years (± SD 0.1) Caucasian 83.4% College level or higher 68.7% Validation sample: 5478 patients from 7 VA medical centers Characteristics: Female 2.7% Mean age 67.8 years (± SD 0.1) Caucasian 83.3% College level or higher 68%	-	-	Comparison: - Results: - Mortality: - All-cause mortality: SIC had a moderate discriminative ability (AUC = 0.71) of SIC in predicting mortality at 2 years' follow-up. A combined model, containing SIC and SF-36 as predictors, had an AUC = 0.74. PROM: - Healthcare utilization: Re-hospitalizations: Discriminative ability of SIC was less able in predicting 2-year re-hospitalizations (AUC = 0.61), which slightly increased when SF-36 was added to the model (AUC = 0.64).
Lorem[50]	All respondents: Angina Asthma Atopic eczema Cancer survivor Cerebrovascular stroke Chronic bronchitis Diabetes Duodenal ulcer Epilepsy Fibromyalgia Food allergies Hand eczema Hypersensitivity Kidney stone Liver disease Migraine Myocardial infarction Osteoporosis Pollen allergies Psoriasis Thyroid Ventricular ulcer For patients >70 years, added: Arthritis Cataract Glaucoma Parkinson disease Rheumatoid arthritis Urinary incontinence	Reference population: 26 684 patients sampled from Tromsø study (1994/1995).[51] Characteristics: Female 52.6% Age group <50 years 61.7% Validation population: 804 patients sampled from Tromsø study and FHI panel (2001/2002). Characteristics: Female 55% Ages 30-79 years	-	Comparison: Medical record-derived CCI Results: HII correlated more strongly with SRH vs CCI (Spearman r = –0.360, $P < .001$ vs r = –0.250, $P < .001$). After excluding all patients with HII = 0, the correlation between HII and SRH strengthened (r = –0.421, $P < .001$) as it weakened between CCI and SRH (r = –0.141, $P < .001$).	Comparison: - Results: Mortality: - PROM: SRH: In an ordinal logistic regression model (containing age, gender, mental health symptoms and the HII), HII had a negative effect (β = –0.249, $P < .001$) after adjustment for the other variables. Healthcare utilization: -
Lucke[52]	For analysis: Asthma Cardiovascular disorder (combined) Coronary heart disease Diabetes mellitus Dyslipidemia GI Hypertension Hyperuricemia Mental disorders Osteoporosis Original questionnaire comprised of 51 (combined) diseases, as well as free text.	2653 patients with COPD or chronic bronchitis. Data from the COSYCONET[53] study. Characteristics: Male 59.4% Mean age 65 years (± SD 8.6) Mean BMI 27 (± SD 5.4) GOLD ≤ 2 57.7%	Comparison: ATC-codes for disease-specific medication Results: Concordance between self-reported comorbidities and ATC codes for disease-specific medication varied from 1.3% (asthma) to 51.8% (combined CVD).	Comparison: Matched ICD-10 codes for diseases and nonspecific medications Results: About 51.5% of self-reported comorbidities were confirmed after comparing them with matched ICD-10 codes.	-

continued on next page

Table 2. Continued

	Items	Study population (n)	Item-level reliability vs other data sources	Overall instrument reliability vs other data sources	Evaluated outcome(s)
Md Yusof[54]	18 sections: 12 physical systems 6 mental health state an additional free text option for medication prescription.	113 community-dwelling patients from 1 research center. Characteristics: Female 56.6% Mean age 75.3 years (± SD 5.19)	-	Comparison: Clinician report Results: CMI correlated significantly with the GP-data (r = 0.8, $P<.001$).	Comparison: - Results: Mortality: Survival: In a Cox proportional hazards model containing 5 continuous predictors (age, weighted CCI, combined condition and age-related CCI score, total score physical sections of CMI, total score mental sections of CMI, and count of medication prescriptions), none of these predictors significantly contributed to predicting mortality. PROM: - Healthcare utilization: -
Merkin[55]	Angioplasty or CABG Cancer Cerebrovascular disease CHF COPD Diabetes Hypertension Myocardial infarction	965 patients with ESRD from 81 dialysis clinics. Data from the CHOICE study.[56] Characteristics: Female 46% Mean age 58 years Caucasian 67% College level or higher 36%	Comparison: Medical record-derived ICED Results: Kappa 0.19 (hypertension) – 0.93 (diabetes) Sensitivity: 18 (COPD) – 96% (diabetes) Specificity: 76 (hypertension) – 98% (diabetes) PPV: - NPV: - Comparison: Clinician report Results: Kappa 0.19 (hypertension) – 0.81 (diabetes) Sensitivity: - Specificity: - PPV: - NPV: -	-	-
Mukerji[57]	Arthritis Diabetes Heart disease Lung disease Other cancers (apart from index tumor) Psychiatric problems Stroke	458 patients with newly diagnosed head-and-neck cancer from 3 hospitals. Characteristics: Female 23.6% Caucasian 86% College level or higher 49.8%	Comparison: Medical record-derived ACE-27 index Results: Kappa: 0.11 (arthritis) – 0.89 (diabetes) Sensitivity: - Specificity: - PPV: - NPV: - Overreporting by self-report occurred more often (21.2%) compared with underreporting (13.5%), with medical records as gold standard.	Comparison: Medical record-derived ACE-27 index Results: Kappa: 0.50 (0.44 – 0.57)	-
Paleri[58]	*9 section headers:* Heart & blood vessels Alcohol consumption Brain and nerves Cancer Diabetes Joints and muscles Kidney Liver, stomach, and pancreas Lungs	20 patients with head-and-neck cancer. Characteristics: -	-	Comparison: Medical record-derived ACE-27 grade Results: Kappa: 0.92 (95% CI 0.82-1.0)	-

continued on next page

Table 2. Continued

	Items	Study population (n)	Item-level reliability vs other data sources	Overall instrument reliability vs other data sources	Evaluated outcome(s)
Pati[59]	Acid peptic disease Arthritis Asthma Cancer Chronic back ache Chronic kidney disease Chronic liver disease (alcohol) Deafness Dementia Diabetes Epilepsy Filariasis Heart disease Hypertension Stroke Thyroid Tuberculosis Visual difficulty	103 patients from 4 primary care practices. Characteristics: Female 45% Mean age 45 years (± SD 5.32)	Comparison: Clinician's prescription-derived data Results: Kappa: 0.58 (hearing problem) – 1.00 (tuberculosis); 16 of 18 conditions were evaluated. Sensitivity: - Specificity: - PPV: - NPV: -	-	-
Lee[60]	Alcohol use Arthritis BMI < 25 Chronic lung disease Current tobacco use Diabetes mellitus Heart failure History of falls History of pain Hypertension Incontinence Memory-related disease Non-skin cancers Other heart problems Psychiatric disease Stroke Visual or hearing impairment	Older (>50 years) community-dwelling patients. Data from the Health and Retirement Study.[61,62] Development cohort: 11 701 patients Characteristics: Female 57% Mean age 67 years (± SD 10) Caucasian 81% High school level or higher 75% Validation Cohort: 8009 patients Characteristics: Female 56% Mean age 67 years (± SD 10) Caucasian 71% High school level or higher 66%	-	-	Results: - Mortality: All-cause mortality: a final model, predicting 4-year mortality, included 12 variables (6 comorbid conditions, sex, age, and 4 functional status measures). Discriminative ability of the model was determined in the development (ROC = 0.84) and validation cohort (ROC = 0.82). - PROM: - - Healthcare utilization: -
Voaklander[63]	Asthma Cancer Chronic back pain Circulatory problems Diabetes Digestive problems Emphysema Epilepsy Eye problems Hay fever/allergies Heart disease Hypertension Kidney disease Liver disease Other Stomach ulcers Stroke Thyroid problems	518 patients receiving major joint arthroplasty at 2 acute care facilities (1995-1997) 283 patients receiving TKA. Characteristics: Female 59% Mean age 69 years (± SD 9) 235 patients receiving THA. Characteristics: Female 60% Mean age 67.1 years (± SD 12)	-	Comparisons: 1. Medical record-derived CCI 2. Administrative data-derived CCI 3. ICD-9 Sum of comorbidities (available in the EHR) Results: SRC correlated significantly with the medical record-derived CCI (Spearman r = 0.40, $P < .01$), administrative data-derived CCI (Spearman r = 0.32, $P < .01$) and ICD-9 Sum (Spearman r = 0.39, $P < .01$).	Comparisons: 1. Medical record-derived CCI 2. Administrative data-derived CCI 3. ICD-9 Sum of comorbidities Results: Mortality: - PROM: SF-36 & WOMAC (health-related quality of life): SRC explained similar amounts in HRQoL domains compared to the ICD-9 Sum, medical record-derived CCI and administrative data-derived CCI. Healthcare utilization: Hospital stay: SRC explained the variance in acute length of stay less better compared to the other comorbidity measures. ER visits: SRC was slightly better in predicting emergency department visits than the other measures (2% variance).

continued on next page

Table 2. Continued

	Items	Study population (n)	Item-level reliability vs other data sources	Overall instrument reliability vs other data sources	Evaluated outcome(s)
Selim[64]	*Physical conditions:* Angina pectoris Cancer Cataract Chronic low back pain Chronic lung disease Congestive heart failure Diabetes Diverticulitis Enlarged prostate Gallbladder disease Gout Heart attack Hepatitis High blood pressure Inflammatory bowel disease Irregular heartbeat Osteoarthritis Osteoporosis Peptic ulcer Peripheral vascular disease Phlebitis Prostatitis Renal failure Rheumatoid arthritis Seizure Skin cancer Stroke Thyroid disease Transient ischemic attacks Urinary tract infection Mental conditions: Alcohol use Anxiety Bipolar disorder Depression Posttraumatic stress disorder Schizophrenia	2425 patients who previously received ambulatory care. Data from Veterans Health Study. Characteristics: Female 0% Mean age 64 years (± SD 12.7) High school level or higher 41%	-	-	Comparison: - Results: Mortality: Survival: In a Cox proportional hazards model, the mortality risk was 14% for every increment in physical CI which decreased to 9% *(P* < .05) after adjustment for sociodemographic and disability rating. PROM: SF-36 (health-related quality of life): Physical CI had the strongest correlations (Pearson's r ≥ 0.29) with SF-36-PCS, while the mental CI correlated better with SF-36-MCS (Pearson's r ≥ 0.30). Both physical and mental CIs were significantly correlated with all SF-36 scales. Combined physical/mental CI also correlated significantly with all HRQoL scales (Pearson's r range –0.29 to –0.45). Healthcare utilization: Hospital visits: Both physical and mental CIs had significant coefficients and explained 5% of the variance in total outpatient clinic visits, which increased to 7% after adjustment for sociodemographic and disability rating.
Vigen[65]	Arthritis Cardiovascular disease Diabetes Gall bladder disease High cholesterol Hypertension Intestinal polyps Irritable bowel syndrome Osteoporosis Other cancers Thyroid disorders	1936 breast cancer patients. Data from the SFBCS[66] & LACE[67] studies. Characteristics: Mean age 60.4 years (± SD 11.0) Caucasian 71% College level or higher 67.8%	Comparison: Medical records (including diagnostic ICD-9 codes) Results: Kappa: 0.50 (other heart diseases) – 0.87 (diabetes) Sensitivity: 48 (other heart diseases) – 90.5% (myocardial infarction) Specificity: 96 (other heart diseases) – 99.5% (diabetes) PPV: - NPV: -	-	Comparison: Medical records (including diagnostic ICD-9 codes) Results: Mortality: All-cause mortality: No significant differences in HRs for covariates between comorbidity models with self-report vs medical records as data source (diabetes HR 1.65 vs 1.44; hypertension 1.22 vs 1.55; myocardial infarction 1.40 vs 1.73; other heart diseases 1.07 vs 1.51) were observed. PROM: - Healthcare utilization: -

- indicates unknown or not described in study

ACE-27 indicates Adult Comorbidity Evaluation – 27; ADL, Activities of Daily Living Scale; ACQUIP, Ambulatory Care Quality Improvement Project; AIDS, acquired immunodeficiency syndrome; ASI, Activity Status Index; ATC, Anatomical Therapeutic Chemical code; AUC, area under the curve; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; BMI, body mass index; BRFSS, Behavior Risk Factor Surveillance System; CABG, coronary artery bypass graft; CAD, coronary artery disease; CCI, Charlson Comorbidity Index; CES-D, Center for Epidemiologic Studies Depression Scale; CHF, congestive heart failure; CHOICE, Choices for Healthy Outcomes in Caring for End-stage renal disease study; CI, confidence interval; CIRS, Cumulative Illness Rating Scale; CMI, Cornell Medical Index; CmSS, Comorbidity Symptom scale; CONNECT, Centralized Nurse-Led Telephone-Based Care Coordination to Improve Outcomes After Surgical Resection for Colorectal Cancer; COPD, chronic obstructive pulmonary disease; CORETH, the Choice of Renal Replacement Therapy project; COSYCONET, COPD and Systemic Consequences – Comorbidities Network Cohort Study; CTD, connective tissue disease; CVA, cerebrovascular accident; CVD, cardiovascular disorder; DBMA, Disease Burden Morbidity Assessment; DMBA-Fv, French version of the Disease Burden Morbidity Assessment; EHR, electronic health record; ELES-PS, Aging in Spain Longitudinal study, Pilot Survey; EQ-5D, EuroQol-5 Dimension; ER, emergency room; ESRD, end-stage renal disease; FHI, Norwegian Institute of Public Health; GHQ-28, General Health Questionnaire – 28; GOLD, Global Initiative for Chronic Obstructive Lung Disease; GP, general practitioner; GSE, general self-efficacy; HAD, Hospital Anxiety and Depression Scale; HII, Health Impact Index; HIV, human immunodeficiency virus; HMO, Health Maintenance Organization; HR, hazard ratio; HRQol, health-related quality-of-life; IBD, inflammatory bowel disease; ICC, intraclass correlation coefficient; ICD-9, International Classification of Diseases version 9; ICD-10, International Classification of Diseases version 10; ICED, Index of Coexistent Disease; LACE, Life After Cancer Epidemiology study; MCS, mental component summary; MMSE, Mini Mental Status Examination; MS, multiple sclerosis; MSC, mental sum scale; mSCQ, modified version of the Self-Administered Comorbidity Questionnaire; NEADL, Nottingham Extended Activities of Daily Living; NP, nurse practitioner; NPV, negative

Table 2. Continued

predictive value; OA, osteoarthritis; OASIS, Outcome in Ankylosing Spondylitis International Study; OR, odds ratio; PA, physician assistant; PAOD, peripheral arterial occlusive disease; PCOS, Prostate Cancer Outcomes Study; PID, pelvic inflammatory disease; PPV, positive predictive value; PROM, patient-reported outcome measure; PSC, physical sum scale; PTCA, percutaneous transluminal coronary angioplasty; PWI, personal wellbeing index; r, Spearman correlation coefficient; RA, rheumatoid arthritis; RCT, randomized controlled trial; ROC, receiver-operating characteristic; Rx, prescription; SCQ, Self-Administered Comorbidity Questionnaire; SCQ-G, German version of the Self-Administered Comorbidity Questionnaire; SD, standard deviation; SF-36-MCS, Short-Form 36 Mental Component Summary scale; SF-36-PCS, Short-Form 36 Physical Component Summary scale; SFBCS, San Francisco Bay Area Cancer Study; SIC, Seattle Index of Co-morbidity; SRC, Self-Reported Co-Morbidity; SRH, Self-Reported Health; THA, total hip arthroplasty; TIA, transient ischemic attack; TKA, total knee arthroplasty; UTI, urinary tract infection; VA, Veterans Affairs; WOMAC, Western Ontario McMaster Osteoarthritis Index; β = βeta.

age, 6 comorbidities, and 4 functional measures) for all-cause mortality at 4 years follow-up.

The most common patient-reported outcome measure evaluated with patient-reported morbidity instruments was the 36-item Short Form Health Survey (SF-36).[24,27,28,32,63,64] One study reported a higher patient-reported CCI score being negatively associated with SF-36 scores,[24] and another reported the number of self-reported conditions in the DBMA instrument being significantly correlated to the SF-36 similar to the medical record–derived CCI.[32] Selim et al[64] also demonstrated a negative association between high CI scores and SF-36 scores. Because of heterogeneous reports on the association and correlation units between both measures in the included studies, there was no clear consistency observed in the direction of the association or correlation.

Several studies also analyzed the relationship between patient-reported comorbidity instruments and healthcare utilization outcomes (eg, [re]hospitalizations, emergency room visits, medical costs).[9,11,26,27,63] Susser et al[9] estimated an AUC of 0.68 and AUC of 0.67 in predicting the number of hospital days and emergency room visits, respectively. Katz et al[11] estimated weak correlations between the patient-reported comorbidity CCI and healthcare utilization outcomes, and Sangha et al[27] reported similar weak associations with the SCQ.

Impact of Demographic Factors on Survey Validity and Reliability

Thirteen studies reported associations between certain patient characteristics and concordance of comorbidity questionnaires with other sources of comorbidity data. Most studies reported higher age being significantly associated with lower agreement between patient-reported and medical-record derived comorbidity data.[22,25,36,39,42,45,47,55,57] Nevertheless, Katz et al,[11] Vigen et al,[65] and Horton et al did not observe a significant association. In terms of reliability, Klabunde et al[5] found a significant association between age ≥ 65 years and inconsistent response patterns between baseline and subsequent surveys. Higher concordance between patient-reported and other comorbidity data sources was also associated with higher education levels.[11,39,42,45] This association was also observed for higher socioeconomic status.[5] In contrast, Vigen et al[65] only reported this association for myocardial infarction. Simpson et al[36] reported that education level did not impact reliability, with Kappa values remaining unaltered after adjustment. Some studies[42,45,55] also reported concordance being significantly influenced by sex, whereas others[44] did not observe a significant observation.

Discussion

Risk adjustment for benchmarking of healthcare outcomes across multiple hospitals is dependent on accurate reporting of case-mix factors such as patient morbidity (ie, comorbid conditions). Clinicians have increasingly considered patient-reported comorbidity instruments as a potential alternative to the laborious review of medical and administrative records or as a method to standardize the collection of data for important conditions across hospitals. Previously published research has looked at multimorbidity measures in primary care or community population settings[68] but, to our knowledge, this is the first systematic literature review focusing on the development or subsequent validation of general patient-reported morbidity instruments (either indices or ad hoc lists of conditions).

Ten original patient-reported comorbidity instruments were found, as well as additional variations on these original studies. The most frequently cited instruments were the SR-CCI and the SCQ, with the SR-CCI demonstrating stronger item-level reliability, overall reliability, and overall agreement, but similar correlations with healthcare utilization parameters. The number of items varied substantially from instrument to instrument. Most studies evaluated the accuracy of self-reported comorbid conditions for individual items compared with another data source, most commonly medical records review. Agreement with the medical record, which was generally used as the gold standard, varied substantially based on the comorbid condition listed within all of the reviewed survey instruments. The kappa values regarding item-level validity were generally not used to eliminate questions with low reliability, leading to large intra-instrument variability. This variation by comorbid condition was thought by the authors to be the result of accurate medical record data with missed diagnoses by the patient and accurate reporting by the patient with missed data in the medical record. The authors postulated that disease items with low reliability included diseases that are "resolved" (in the past),[65] those that are controlled with treatment (eg, hypertension),[65] those without symptoms,[36] those with complex diagnostic criteria or ambiguous disease categories (eg, heart diseases such as atherosclerosis or heart failure),[24,65] and those with confusing or overlapping names (eg, arthritis vs osteoarthritis vs rheumatoid arthritis).[24,36,44,57] Violán et al[69] demonstrated similar results in a cross-sectional study comparing morbidity prevalence between electronic health records and health surveys; self-reported morbidity prevalence was higher among younger patients and for symptomatic conditions. Diseases with clear definitions (eg, diabetes) and that required ongoing treatment had higher agreement with other data sources and were most accurately reported by patients.[25,36,45] Even in those cases, there may have been disagreement, such as if people with pre-diabetes classified themselves as diabetic or people with non–insulin-dependent diabetes considered themselves not diabetic. Agreement between patient-reported morbidity instruments and administrative data was usually poor, with the limitation generally listed in some studies that administrative data may underreport the presence of comorbid conditions ("undercoding") more than the medical record.

This systematic review highlights the lack of information on the predictive validity of comorbidity data; a subset of included studies examined the predictive capabilities of the patient-reported comorbidity questionnaires/indices for outcomes such as mortality, patient-reported outcome measures (such as functional status or general health-related quality of life), and healthcare utilization and costs. The patterns of correlations between patient-reported comorbidity data and mortality were as

hypothesized: higher patient-reported comorbidity scores were associated with poorer overall survival.[25] The predictive ability of patient-reported comorbidity indices were moderate to good (AUCs > 0.70) for all-cause mortality, regardless of whether only comorbidities were used or other variables were added to the model.[26,48,60] Unsurprisingly, higher patient-reported comorbidity scores were significantly correlated and associated with lower health-related quality of life scores (as measured by the SF-36).[24,27,28,32,63,64] Patient-reported comorbidity scores had low positive (Spearmans r < 0.50) correlations with many healthcare utilization measures (eg, hospitalizations, length of hospital stay, prescribed medications)[11,27] and had poor discriminative ability (AUCs between 0.60-0.70) for these measures as well.[9,48] Because comorbidity measures can influence outcomes and interpretation when comparing treatment strategies or hospitals, it is essential that a comorbidity measure is validated for the population *and* outcome of interest.

Gold standard survey instrument development generally begins with a systematic exploration of potential topics to include in survey questions, followed by the development of a large number of questions, which overlap and then reduce to a smaller number of higher performing questions based on field testing.[70,71] For example, a question on the presence or absence of myocardial infarction could be asked several ways, and the question that proves most valid and reliable in testing is retained and the other forms of the question are discarded. Although 7 studies mentioned that their questionnaires were rephrased to optimize clarity and comprehension,[27,32,33,38,44,47,59] there were no studies that started intentionally with a larger question bank that was then reduced down to reliable and valid questions.

This study should be interpreted in the context of its limitations. Although an extensive search in 5 electronic database was conducted with the help of an experienced librarian, the possibility of missing studies cannot be excluded. Comorbidity data is collected in a large proportion of medical studies, and therefore inclusion and exclusion criteria had to be defined to allow for reasonable identification of self-reported comorbidity instruments. The goal was to identify studies that primarily focused on instrument development or validation. It is possible that comorbidity questionnaires that were used but did not have any description of their development, validity, or reliability in the abstract or title could have been missed in the present search strategy, such as instruments in the gray literature. Additionally, studies that may have subsequently used information from a comorbidity instrument to predict an outcome in a general study without mention of the instrument itself in the abstract could have been missed with this search strategy. The presence of 2 independent reviewers to conduct the study selection, data extraction, and overall interpretation added to the accuracy of the review process.

This systematic review was inspired by international efforts (eg, the International Consortium for Health Outcomes Measurement [ICHOM],[3] the United States Agency for Healthcare Research and Quality (AHRQ),[72] and the Registry of Patient Registries project[73]) to harmonize collection and format of outcome measures. These outcome measures can include suggested outcome domains, measurement tools, and predictor variables for risk adjustment for a given medical condition. Among these risk-adjustment factors are morbidity variables, which can be collected and analyzed using various methods. Minimal standards for consistent capture of morbidity data are essential for fair benchmarking for VBHC or public reporting or outcomes. Patient-reported morbidity instruments can be used internationally and among hospitals with no or different electronic medical records and administrative data structure. Additionally, these instruments could be used to bolster the medical record. Future research should focus on the capture of complete morbidity data for the purposes of more robust risk adjustment, a key component of fair benchmarking for VBHC.

Conclusions

In summary, this systematic review found ten self-reported morbidity instruments, with the Patient-Reported Charlson Comorbidity Index[11] and the Self-Administered Comorbidity Questionnaire[27] being the most frequently cited instruments. Within each included instrument, there was significant variability in the reliability of patient-reported comorbidities based on the comorbid condition. Further research is needed to determine whether patient-reported comorbidity data should be used to bolster medical records data or serve as a stand-alone entity for risk adjustment of observational outcomes data.

Acknowledgments

The authors would like to acknowledge Wichor M. Bramer (Biomedical Information Specialist, Medical Library, Erasmus MC) for his support in the literature search.

Authors A. Oemrawsingh and J.A. Hazelzet were supported by a grant from the Federation of Dutch University Medical Centers (NFU). R.W. Bergmark received grants from the Gliklich Healthcare Innovation Scholar Fund/Massachusetts Eye and Ear Infirmary and from the American Board of Medical Specialties during the development of this paper.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2020.02.006.

REFERENCES

1. Fortin M, Soubhi H, Hudon C, Bayliss EA, van den Akker M. Multimorbidity's many challenges. *BMJ*. 2007;334(7602):1016–1017.
2. Valderas JM, Starfield B, Sibbald B, Salisbury C, Roland M. Defining comorbidity: implications for understanding health and health services. *Ann Fam Med*. 2009;7(4):357–363.
3. Obbarius A, van Maasakkers L, Baer L, et al. Standardization of health outcomes assessment for depression and anxiety: recommendations from the ICHOM Depression and Anxiety Working Group. *Qual Life Res*. 2017;26(12):3211–3225.
4. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8–27.
5. Klabunde CN, Harlan LC, Warren JL. Data sources for measuring comorbidity: a comparison of hospital records and Medicare claims for cancer patients. *Med Care*. 2006;44(10):921–928.
6. Klabunde CN, Warren JL, Legler JM. Assessing comorbidity using claims data: an overview. *Med Care*. 2002;40(suppl 8):IV-26-35.
7. de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity: a critical review of available methods. *J Clin Epidemiol*. 2003;56(3):221–229.
8. Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T. Comorbidities, complications, and coding bias. Does the number of diagnosis codes matter in predicting in-hospital mortality? *JAMA*. 1992;267(16):2197–2203.
9. Susser SR, McCusker J, Belzile E. Comorbidity information in older patients at an emergency visit: self-report vs. administrative data had poor agreement but similar predictive validity. *J Clin Epidemiol*. 2008;61(5):511–515.
10. Preen DB, Holman CDAJ, Lawrence DM, Baynham NJ, Semmens JB. Hospital chart review provided more accurate comorbidity information than data from a general practitioner survey or an administrative database. *J Clin Epidemiol*. 2004;57(12):1295–1304.
11. Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW. Can comorbidity be measured by questionnaire rather than medical record review? *Med Care*. 1996;34(1):73–84.
12. Olomu AB, Corser WD, Stommel M, Xie Y, Holmes-Rovner M. Do self-report and medical record comorbidity data predict longitudinal functional capacity and quality of life health outcomes similarly? *BMC Health Serv Res*. 2012;12(1):398.
13. Katherine F, Amy T, Nicole B, Sigrid C, Andrew JV. Comparison of physician-documented versus patient-reported collection of comorbidities among

patients with prostate cancer upon first visit to the urology clinic. *JCO Clin Cancer Inform*. 2018;(2):1–10.
14. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
15. Bramer WM, Rethlefsen ML, Mast F, Kleijnen J. Evaluation of a new method for librarian-mediated literature searches for systematic reviews. *Res Synth Methods*. 2018;9(4):510–520.
16. Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Syst Rev*. 2017;6(1):245.
17. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. *J Med Libr Assoc*. 2016;104(3):240–243.
18. Altman DG. *Practical Statistics for Medical Research*. Boca Raton, FL: CRC Press; 1990.
19. Akoglu H. User's guide to correlation coefficients. *Turk J Emerg Med*. 2018;18(3):91–93.
20. Gliklich RE, Leavy MB, Karl J, Campion DM, Levy D, Berliner E. A framework for creating standardized outcome measures for patient registries. *J Comp Eff Res*. 2014;3(5):473–480.
21. McCusker J, Verdon J, Tousignant P, de Courval LP, Dendukuri N, Belzile E. Rapid emergency department intervention for older people reduces risk of functional decline: results of a multicenter randomized trial. *J Am Geriatr Soc*. 2001;49(10):1272–1281.
22. Corser W, Sikorskii A, Olomu A, Stommel M, Proden C, Holmes-Rovner M. Concordance between comorbidity data from patient self-report interviews and medical record documentation. *BMC Health Serv Res*. 2008;8:85.
23. Holmes-Rovner M, Stommel M, Corser WD, et al. Does outpatient telephone coaching add to hospital quality improvement following hospitalization for acute coronary syndrome? *J Gen Intern Med*. 2008;23(9):1464–1470.
24. Ng X, Low AH, Thumboo J. Comparison of the Charlson Comorbidity Index derived from self-report and medical record review in Asian patients with rheumatic diseases. *Rheumatol Int*. 2015;35(12):2005–2011.
25. Habbous S, Chu KP, Harland LT, et al. Validation of a one-page patient-reported Charlson comorbidity index questionnaire for upper aerodigestive tract cancer patients. *Oral Oncol*. 2013;49(5):407–412.
26. Chaudhry S, Jin L, Meltzer D. Use of a self-report-generated Charlson Comorbidity Index for predicting mortality. *Med Care*. 2005;43(6):607–615.
27. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The self-administered comorbidity questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Rheum*. 2003;49(2):156–163.
28. Stolwijk C, van Tubergen A, Ramiro S, et al. Aspects of validity of the self-administered comorbidity questionnaire in patients with ankylosing spondylitis. *Rheumatology (Oxford)*. 2014;53(6):1054–1064.
29. Spoorenberg A, van der Heijde D, de Klerk E, et al. Relative value of erythrocyte sedimentation rate and C-reactive protein in assessment of disease activity in ankylosing spondylitis. *J Rheumatol*. 1999;26(4):980–984.
30. Robinski M, Strich F, Mau W, Girndt M. Validating a patient-reported comorbidity measure with respect to quality of life in end-stage renal disease. *PLoS One*. 2016;11(6):e0157506.
31. Robinski M, Mau W, Lamprecht J, Krauth C, Girndt M. The Choice of Renal Replacement Therapy (CORETH) project: study design and methods. *Clin Kidney J*. 2014;7(6):575–581.
32. Bayliss EA, Ellis JL, Steiner JF. Subjective assessments of comorbidity correlate with quality of life health outcomes: initial validation of a comorbidity assessment instrument. *Health Qual Life Outcomes*. 2005;3:51.
33. Poitras ME, Fortin M, Hudon C, Haggerty J, Almirall J. Validation of the disease burden morbidity assessment by self-report in a French-speaking population. *BMC Health Serv Res*. 2012;12:35.
34. Wijers IGM, Ayala A, Rodriguez-Blazquez C, Rodriguez-Laso A, Rodriguez-Rodriguez V, Forjaz MJ. Disease burden morbidity assessment by self-report: psychometric properties in older adults in Spain. *Geriatr Gerontol Int*. 2017;17(7):1102–1108.
35. Teófilo Rodríguez J, González Cabezas AN, Díaz Veiga P, Rodríguez Rodríguez V. Estudio Longitudinal Envejecer en España: El proyecto ELES. *Perfiles y Tendencias*. 2011;50:1–44.
36. Simpson CF, Boyd CM, Carlson MC, Griswold ME, Guralnik JM, Fried LP. Agreement between self-report of disease diagnoses and medical record validation in disabled older women: factors that modify agreement. *J Am Geriatr Soc*. 2004;52(1):123–127.
37. Kasper JD, Shapiro S, Guralnik JM, Bandeen-Roche KJ, Fried LP. Designing a community study of moderately to severely disabled older women: the Women's Health and Aging Study. *Ann Epidemiol*. 1999;9(8):498–507.
38. Crabtree HL, Gray CS, Hildreth AJ, O'Connell JE, Brown J. The comorbidity symptom scale: a combined disease inventory and assessment of symptom severity. *J Am Geriatr Soc*. 2000;48(12):1674–1678.
39. De-loyde KJ, Harrison JD, Durcinoska I, Shepherd HL, Solomon MJ, Young JM. Which information source is best? Concordance between patient report, clinician report and medical records of patient co-morbidity and adjuvant therapy health information. *J Eval Clin Pract*. 2015;21(2):339–346.
40. Young JM, Butow PN, Walsh J, et al. Multicenter randomized trial of centralized nurse-led telephone-based care coordination to improve outcomes after surgical resection for colorectal cancer: the CONNECT intervention. *J Clin Oncol*. 2013;31(28):3585–3591.
41. Gad BV, Higuera CA, Klika AK, Elsharkawy KA, Barsoum WK. Validity of patient-reported comorbidities before total knee and hip arthroplasty in patients older than 65 years. *J Arthroplasty*. 2012;27(10):1750–1756.e1s.
42. Hansen H, Schafer I, Schon G, et al. Agreement between self-reported and general practitioner-reported chronic conditions among multimorbid patients in primary care - results of the MultiCare Cohort Study. *BMC Fam Pract*. 2014;15:39.
43. Schafer I, Hansen H, Schon G, et al. The German MultiCare-study: patterns of multimorbidity in primary health care - protocol of a prospective cohort study. *BMC Health Serv Res*. 2009;9:145.
44. Horton M, Rudick RA, Hara-Cleaver C, Marrie RA. Validation of a self-report comorbidity questionnaire for multiple sclerosis. *Neuroepidemiology*. 2010;35(2):83–90.
45. Iecovich E, Biderman A. Concordance between self-reported and physician-reported chronic co-morbidity among disabled older adults. *Can J Aging*. 2013;32(3):287–297.
46. Potosky AL, Harlan LC, Stanford JL, et al. Prostate cancer practice patterns and quality of life: the Prostate Cancer Outcomes Study. *J Natl Cancer Inst*. 1999;91(20):1719–1724.
47. Boissonnault WG, Badke MB. Collecting health history information: the accuracy of a patient self-administered questionnaire in an orthopedic outpatient setting. *Phys Ther*. 2005;85(6):531–543.
48. Fan VS, Au D, Heagerty P, Deyo RA, McDonell MB, Fihn SD. Validation of case-mix measures derived from self-reports of diagnoses and health. *J Clin Epidemiol*. 2002;55(4):371–380.
49. Fihn SD, McDonell MB, Diehr P, et al. Effects of sustained audit/feedback on self-reported health status of primary care patients. *Am J Med*. 2004;116(4):241–248.
50. Lorem GF, Schirmer H, Emaus N. Health Impact Index. Development and validation of a method for classifying comorbid disease measured against self-reported health. *PLoS One*. 2016;11(2):e0148830.
51. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njolstad I. Cohort profile: the Tromso Study. *Int J Epidemiol*. 2012;41(4):961–967.
52. Lucke T, Herrera R, Wacker M, et al. Systematic analysis of self-reported comorbidities in large cohort studies - a novel stepwise approach by evaluation of medication. *PLoS One*. 2016;11(10):e0163408.
53. Karch A, Vogelmeier C, Welte T, et al. The German COPD cohort COSYCONET: aims, methods and descriptive analysis of the study population at baseline. *Respir Med*. 2016;114:27–37.
54. Md Yusof MY, Horan MA, Jones M, McInnes L, Rabbitt PM, Pendleton N. Developing a self-reported comorbidity index to predict mortality of community-dwelling older adults. *Arch Gerontol Geriatr*. 2010;50(3):e63–67.
55. Merkin SS, Cavanaugh K, Longenecker JC, Fink NE, Levey AS, Powe NR. Agreement of self-reported comorbid conditions with medical and physician reports varied by disease among end-stage renal disease patients. *J Clin Epidemiol*. 2007;60(6):634–642.
56. Powe NR, Klag MJ, Sadler JH, et al. Choices for healthy outcomes in caring for end stage renal disease. *Seminars in Dialysis*. 1996;9(1):9–11.
57. Mukerji SS, Duffy SA, Fowler KE, Khan M, Ronis DL, Terrell JE. Comorbidities in head and neck cancer: agreement between self-report and chart review. *Otolaryngol Head Neck Surg*. 2007;136(4):536–542.
58. Paleri V, Wight RG. A cross-comparison of retrospective notes extraction and combined notes extraction and patient interview in the completion of a comorbidity index (ACE-27) in a cohort of United Kingdom patients with head and neck cancer. *J Laryngol Otol*. 2002;116(11):937–941.
59. Pati S, Hussain MA, Swain S, et al. Development and validation of a questionnaire to assess multimorbidity in primary care: an Indian experience. *Biomed Res Int*. 2016;2016:6582487.
60. Lee SJ, Lindquist K, Segal MR, Covinsky KE. Development and validation of a prognostic index for 4-year mortality in older adults. *JAMA*. 2006;295(7):801–808.
61. The Health and Retirement Survey. Ann Arbor: University of Michigan; 1996. http: //hrsonline.isr.umich.edu/docs/sample/sho_samp.php?hfyle=ref023b&xtyp=1. Accessed January 19, 2006.
62. The Health and Retirement Survey. Ann Arbor: University of Michigan. http://hrsonline.isr.umich.edu/intro/sho_uinfo.php?hfyle=overview&xtyp=2. Accessed January 19, 2006.
63. Voaklander DC, Kelly KD, Jones CA, Suarez-Almazor ME. Self report co-morbidity and health related quality of life – a comparison with record based co-morbidity measures. *Social Indicators Research*. 2004;66(3):213–228.
64. Selim AJ, Fincke G, Ren XS, et al. Comorbidity assessments based on patient report: results from the Veterans Health Study. *J Ambul Care Manage*. 2004;27(3):281–295.
65. Vigen C, Kwan ML, John EM, et al. Validation of self-reported comorbidity status of breast cancer patients with medical records: the California Breast Cancer Survivorship Consortium (CBCSC). *Cancer Causes Control*. 2016;27(3):391–401.
66. John EM, Phipps AI, Davis A, Koo J. Migration history, acculturation, and breast cancer risk in Hispanic women. *Cancer Epidemiol Biomarkers Prev*. 2005;14(12):2905–2913.

67. Caan B, Sternfeld B, Gunderson E, Coates A, Quesenberry C, Slattery ML. Life After Cancer Epidemiology (LACE) Study: a cohort of early stage breast cancer survivors (United States). *Cancer Causes Control*. 2005;16(5):545–556.
68. Huntley AL, Johnson R, Purdy S, Valderas JM, Salisbury C. Measures of multimorbidity and morbidity burden for use in primary care and community settings: a systematic review and guide. *Ann Fam Med*. 2012;10(2):134–141.
69. Violán C, Foguet-Boreu Q, Hermosilla-Pérez E, et al. Comparison of the information provided by electronic health records data and a population health survey to estimate prevalence of selected health conditions and multimorbidity. *BMC Public Health*. 2013;13(1):251.
70. Sprangers MA, Cull A, Bjordal K, Groenvold M, Aaronson NK. The European Organization for Research and Treatment of Cancer. Approach to quality of life assessment: guidelines for developing questionnaire modules. EORTC Study Group on Quality of Life. *Qual Life Res*. 1993;2(4):287–295.
71. Farnik M, Pierzchała WA. Instrument development and evaluation for patient-related outcomes assessments. *Patient Relat Outcome Meas*. 2012;3:1–7.
72. Leroy L, Bayliss E, Domino M, et al. The Agency for Healthcare Research and Quality Multiple Chronic Conditions Research Network: overview of research contributions and future priorities. *Med Care*. 2014;52(Suppl 3):S15–S22.
73. Gliklich RE DN, Leavy MB, eds. *Registries for evaluating patient outcomes: a user's guide [Internet]*. 3rd ed. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 Apr. 1. Patient Registries. https://www.ncbi.nlm.nih.gov/books/NBK208643/. Accessed July 24, 2019.