

## How to measure comorbidity: a critical review of available methods

Vincent de Groot<sup>a,b,\*</sup>, Heleen Beckerman<sup>a,b</sup>, Gustaaf J. Lankhorst<sup>a,b</sup>, Lex M. Bouter<sup>b</sup>

<sup>a</sup>Department of Rehabilitation Medicine, VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

<sup>b</sup>Institute for Research in Extramural Medicine, VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

Received 15 February 2001; received in revised form 19 July 2002; accepted 27 August 2002

### Abstract

The object of this article was to systematically review available methods to measure comorbidity and to assess their validity and reliability. A search was made in Medline and Embase, with the keywords comorbidity and multi-morbidity, to identify articles in which a method to measure comorbidity was described. The references of these articles were also checked, and using a standardized checklist the relevant data were extracted from these articles. An assessment was made of the content, concurrent, predictive and construct validity, and the reliability. Thirteen different methods to measure comorbidity were identified: one disease count and 12 indexes. Data on content and predictive validity were available for all measures, while data on construct validity were available for nine methods, data on concurrent validity, and interrater reliability for eight methods, and data on intrarater reliability for three methods. The Charlson Index is the most extensively studied comorbidity index for predicting mortality. The Cumulative Illness Rating Scale (CIRS) addresses all relevant body systems without using specific diagnoses. The Index of Coexisting Disease (ICED) has a two-dimensional structure, measuring disease severity and disability, which can be useful when mortality and disability are the outcomes of interest. The Kaplan Index was specifically developed for use in diabetes research. The Charlson Index, the CIRS, the ICED and the Kaplan Index are valid and reliable methods to measure comorbidity that can be used in clinical research. For the other indexes, insufficient data on the clinimetric properties are available. © 2003 Elsevier Science Inc. All rights reserved.

*Keywords:* Comorbidity; Multimorbidity; Measurement; Reliability; Validity; Prediction; Review

### 1. Introduction

As early as 1970, Alvan Feinstein noted that “the failure to classify and analyze comorbid diseases has led to many difficulties in medical statistics” [1], because comorbidity can affect the moment of detection, prognosis, therapy, and outcome. Comorbidity can play an important role in different types of research. In etiologic studies the relationship between comorbid conditions and an index disease can be investigated. Comorbidity can be the cause or the consequence of an index disease. It is also possible that the index disease and the comorbid conditions share the same risk factors. In diagnostic studies, comorbidity can obscure the relationship between the test under study and the index disease. In these fields of research it might be particularly useful to analyze every disease as a separate variable, to gain insight into the relationship between individual diseases and the index disease at issue. However, this method is not feasible in small studies, because of reduced efficiency of the

analysis. Randomized controlled trials (RCTs) and prognostic studies can also be complicated by comorbidity. Comorbidity can either act as a confounder, threatening the internal validity, or as an effect modifier, threatening the internal and external validity of the study. For these purposes an efficient method is needed to measure comorbidity.

There are four important reasons for measuring comorbidity. The first reason is to be able to correct for confounding, and thus improve the internal validity of studies. The second reason is to be able to identify effect modification. A third reason is the desire to use comorbidity as a predictor of study outcome or natural history. Finally, a comprehensive comorbidity measure, including many cooccurring comorbid conditions in one valid variable, is needed for reasons of statistical efficiency.

Because an overview of available methods to measure comorbidity is still lacking, the following research question was formulated: Which methods are available for measuring comorbidity that can be used in RCTs and prognostic studies?

### 2. Methods

A search was made in the electronic databases of Medline (from January 1966 to September 2000) and Embase

\* Corresponding author. Tel.: +31 20-4440756; fax: +31 20-4440787.

E-mail address: v.degroot@vumc.nl (V. de Groot).

(from January 1988 to September 2000). The following keywords were used to identify potentially useful articles: comorbidity, multimorbidity, and coexisting disease. Articles in which the focus was on comorbidity assessment or comorbidity was an important (prognostic) variable were considered for inclusion. In the literature several terms are being used in comorbidity research. Also, there is no consensus regarding the definition of these terms. This review focuses on methods that can be used to assess the burden of diseases that exists besides an index disease, i.e., the disease under study. Subsequently, it was carefully checked whether the articles described methods to assess comorbidity or the clinimetric properties of these methods. Then, the reference lists of the retrieved articles were checked for other eligible articles. Methods that analyze every disease as a separate variable were excluded. Because we focus on comorbidity in medical patients, methods that deal with psychiatric comorbidity in populations in which the index disease (the main disease under study) is also psychiatric were also excluded.

An assessment was made of the content, criterion, and construct validity, as well as the reliability of the identified methods [2]. As an indication of the administrative burden a description of the information that is needed to arrive at a score on the measure is given. Because the focus of this review is on comorbidity as a determinant and not on comorbidity as an evaluative measure, responsiveness was not assessed. Content validity concerns the extent to which a measure includes all relevant items: it is a qualitative assessment. To describe content validity a short description is given of the items included in the method, whether or not some type of weighting or (pathophysiologic) severity ranking was applied, which information is needed to obtain a score, how to arrive at the final score, and whether or not adaptations of the method are available for specific purposes.

Streiner and Norman [2] define criterion validity as: “the correlation of a scale with some other measure of the trait or disorder under study, ideally, a ‘gold standard’ (the criterion) which has been used and accepted in the field.” Unfortunately, there is no “gold standard” available for measuring comorbidity in medical patients, so one has to use other comorbidity measures for comparison. In this situation the decision on which measure is best depends not only on statistical tests but also on clinical judgement. Criterion validity can be subdivided into concurrent and predictive validity. Concurrent validity is assessed by correlating the measure under study with the criterion measure, which is given at the same time [2]. Parameters used to assess concurrent validity are the Spearman or Pearson correlation coefficients ( $r$ ) and the Intraclass Correlation Coefficient (ICC). Although it is very difficult to determine cutoff points for correlation coefficients, because there are many factors influencing their value [2], correlation coefficients exceeding 0.40 were considered to be moderate and those exceeding 0.75 were considered to be high [3]. Predictive validity [2] is the ability of a measure to predict future events or future scores on the

outcome measure of interest. The assessment of predictive validity was based on parameters obtained from survival analysis, proportional hazards models, and linear or logistic regression models. Points of interest were the relative risks (RR), relative hazards (RH), odds ratios (OR), explained variance ( $r^2$ ), and the area under the receiver operating characteristic curve (AUC). If regression models predicting future events were significant or significantly improved after adding the comorbidity measure under study, this was considered to support predictive validity.

The assessment of construct validity encompasses the testing of hypotheses regarding the relationship of the measure under study with other more or less related traits (constructs) [2], such as age, mortality, ADL, length of stay, or number of medications taken. There are several methods that can be used to assess construct validity, such as correlation coefficients and comparing means or proportions in different populations. Whether or not construct validity is confirmed will be discussed for every measure, because it is not possible to formulate comprehensive rules for assessing construct validity.

Different types of reliability were also investigated: test-retest reliability, and intra- and interrater reliability. Parameters used to assess reliability are (in descending order of appropriateness) [2] Intraclass Correlation Coefficients (ICC), (weighted) Cohens Kappa [(w)K], correlation coefficients ( $r$ ), and percentage of reliability. Reliability coefficients are considered to be fair when they exceed 0.40 and moderate when they exceed 0.75 [3].

For every identified method the first author (VdG) screened all related articles for data regarding the clinimetric properties of that method, using a standardized data-extraction form. Every method was either classified as an “index” or as a “disease count.” Methods were classified as an “index” if the authors used weights or (pathophysiologic) severity rankings for the conditions or dimensions included in the index. Methods were classified as a “disease count” if the authors solely used an enumeration of the number of conditions present. Methods that did not already have a name were given the name of the first author and the category to which they were assigned. For example, if the first author’s name is Schwarz, and the method was classified as an index, the method is referred to as the Schwarz index.

### 3. Results

Thirteen different methods to assess comorbidity were identified and presented in alphabetical order in Table 1: one disease count and 12 indexes. Data on content and predictive validity were available for all measures, while data on construct validity were available for nine methods, data on concurrent validity for eight methods, data on interrater reproducibility for eight methods, and data on intrarater reliability for three methods.

The Burden of Disease (BOD) index [4] consists of 59 weighted disease categories, selected on the basis of a litera-

Table 1  
Comorbidity measures: main characteristics and study populations

Index	Items	Weights	Information needed	Final score	Adaptations	Study populations (n)
BOD index	59 diseases	0—Not present 1—Inactive 2—Mild 3—Moderate 4—Severe 0—RR 1.2–1.5 1—RR 1.5–2.5 2—RR 2.5–3.5 3—RR 3.5–4.5 6—RR > 6	Clinical assessment of symptoms, complications, need for and complexity of therapy	Sum of weights	None	Long-stay nursing home patients (194)
Charlson index	19 conditions	0—No impairment 4—Life-threatening impairment	Clinical judgment	Sum of weights	—ICD-9 —Age —Patients with amputations —Questionnaire	Breast cancer (685), HIV+ (129), ICU (201), SCI (330), Stroke (106), Cancer (203), Several surgical procedures (>10,000), Heart disease (>10,000), Pneumonia (>10,000), Elective noncardiac surgery with diabetes or hypertension (218), Lower limb amputees (24)
CURS	13 body systems	0—No impairment 4—Life-threatening impairment	Clinical judgment	Sum of weights	None	Mixed inpatients (472), Mixed deceased patients (72), Elderly and geriatric outpatients (141 + 181), Mixed institutionalized long-term care patients (439), SCI (330), Cancer (203)
Comroni-Huntley index		1—No comorbidity 2—Impaired vision or hearing 3—Heart disease, stroke or diabetes 4—Both levels 2 and 3 None	Not specified	1–4	None	Hypertension and age 75 to 84 years (878)
Disease count	Single diseases		—Interview —Patient record —ICD-9 codes	Number of present diseases, maximal score depends on the limits that are applied	Several	SCI (330), HIV+ (395), ICU (105), Breast cancer (>10,000), Myocardial infarction (>10,000), Asthma (>10,000), Appendicitis (>10,000), Abdominal hernia (>10,000), Diverticulitis (>10,000), Biliary tract disease (>10,000), Low back pain (>10,000), Pneumonia (>10,000), Diabetes with complications (>10,000), Rehabilitation inpatients (>10,000)
DUSOI index	Every present health problem is rated on four domains: —Symptom level —Complication level —Prognosis without treatment —Treatability (= prognosis with treatment)	0–5	Clinical judgment	Using a weighted scoring paradigm leading to a score 0–100	None	Mixed primary care patients (414+1191)

(continued)

Table 1  
(Continued)

Index	Items	Weights	Information needed	Final score	Adaptations	Study populations (n)
Hallstrom index	2 domains: —CF consisting of 10 conditions —SF consisting of 6 cardiac symptoms —No comorbidity —Non-disabling comorbidity —Disabling comorbidity	None	Interview	—CF number of present conditions —SF number of present symptoms —Total $1.67 \times CF + SF$ None	None	Out of hospital ventricular fibrillation (282)
Hurwitz index			Not specified		None	Back-related problems (931)
ICED						
DS	—14 disease categories	—1–5	—Symptoms, signs and laboratory tests	Using a scoring paradigm leading to scores 1–4	None	Total hip replacement (356), Long-stay nursing home residents (194)
FS	—10 functional areas	—1–3	—Level of impairment	Sum of weights	Adding points for every decade over 75	Mixed geriatric and general medicine (370)
Incalzi index	52 Conditions	Based on RR for mortality	—History —Physical examination —Routine laboratory data —ECG			
Kaplan index	Vascular or non-vascular disease	0—Noncogent, easy to control or no comorbidity 1—Slight decompensation of vital system or nonthreatening chronic conditions 2—Impaired vital system or potentially threatening chronic condition 3—Recent full decompensation of vital system or life-threatening chronic condition	—Chest X-ray Clinical information	According to the most severe condition, two grades 2 are ranked as 3	Expanded with several diagnoses	Diabetes (188), Breast cancer (404)
Liu index	38 conditions	0—Not present 5—Active rehabilitation contraindicated	Medical records	Sum of weights	None	Stroke (106)
Shwartz index	21 conditions	Regression coefficient from a model to predict costs	Medical records of databases with ICD-9 codes	Sum of weights of present conditions	None	Mixed patients (4439); Stroke, Lung disease, Heart disease, Prostate disease, Hip and femur fracture, low back disorder

BOD, Burden of Disease index; CF, chronic factor; CIRRS, Cumulative Illness Rating Scale; DS, disease severity; DUSOI, Duke Severity of Illness; ECG, electrocardiogram; FS, functional severity; HIW +, human immunodeficiency virus positive; ICD-9, International Classification of Diseases, version 9; ICED, Index of Coexistent Disease; ICU, intensive care unit; RR, relative risk; SCI, spinal cord injury; SF, symptom factor.

For more detailed information on the validity and reliability of the comorbidity measures please visit our Website <http://www.emgo.nl/publications/> or <http://www.vumc.nl/revalidatie/index.html> and click on English information, downloads, comorbidity indexes.

ture review and a consensus meeting of three physicians and a nurse. To obtain the score on the BOD Index, medical records over the previous year were reviewed, based on standardized guidelines assessing symptoms, complications and need for and complexity of therapy. Although the clinical data are obtained from one single article, they support the concurrent and predictive validity of the BOD [4]. Furthermore, the authors found a low, but positive relationship with the Katz Activities of Daily Living scale (Katz ADL), supporting construct validity [4]. There was no relationship with the Sickness Impact Profile (SIP), whereas a weak positive relationship would be expected. Interrater reliability was good [4] (Table 1).

The Charlson Index [5] is the most extensively studied comorbidity index [4–26]. The 19 diseases included in the index have been selected and weighted on the basis of the strength of their association with mortality. It has been adapted for use with ICD-9 databases [7–13,15,22], for use with patients with amputations [17], transformed into a questionnaire [6], and combined with age to form an age-comorbidity index [16]. Four out of six comparisons with other indices of comorbidity yielded correlation coefficients exceeding 0.40, supporting concurrent validity [4,18–22]. Predictive validity was confirmed by finding many significant relationships of the Charlson index with various criterion outcomes, such as mortality, disability, readmissions and length of stay (Table 1) [5,7,10–14,16,17,19,22–25]. All relationships with various kinds of variables showed some, although not perfect, correlations in the anticipated directions, supporting construct validity [4,6,8,9,12,14,15,18–20,24,26]. Test-retest reliability is good, and interrater reliability is moderate to good (except for one outlier with an ICC = 0.16) [6,14,18,20]. It should be noted that emphasis has been laid on the ability of the index to predict mortality (Table 1).

The Cumulative Illness Rating Scale (CIRS) [27] rates 13, conceptually valid, body systems (supporting content validity) on a five-point (pathophysiologic) severity scale. It has been slightly adapted to form the CIRS-G (CIRS geriatric) [28], for which guidelines to enhance reliability have been formulated. Criterion validity has been confirmed by showing high correlation coefficients when comparing CIRS scores based on autopsy (the gold standard) with those based on health histories and chart reviews [29]. The CIRS was correlated with four other measures of comorbidity. Three out of five correlation coefficients exceeded 0.40, supporting concurrent validity [19,20,28]. There is little evidence to support predictive validity [19,30]. Small to fair positive correlations in the anticipated directions have been found for other variables, such as medication usage, ADL, IADL, and age, supporting construct validity [20,28,31]. Interrater and test-retest reliability are good [19,20,27,28] (Table 1).

The Cornoni-Huntley index was intended to be used in a study investigating hypertension and associated comorbid conditions [32]. Because the authors used data that were gathered in another study, only information on visual acuity, hearing ability, heart disease, stroke, and diabetes was

available. Based on this information, they constructed a four-level comorbidity index. The limited data support the predictive validity with mortality as outcome [32] (Table 1).

Several authors studied comorbidity by simply counting the number of diseases that exist in addition to the index disease of a patient [19,33–36]. Although this method seems to be quite straightforward, substantial differences exist with regard to the definitions used to define a condition as comorbid. Some authors used ICD-9 codes to count the total number of comorbid conditions [19], whereas others made up a list of carefully selected comorbid conditions and counted the number of these conditions that were present, using medical records or ICD-9-CM codes [33–35]. Gross et al. [36] defined a condition as being comorbid if it required treatment or had altered an organ function. Three out of five correlations with other comorbidity measures or severity of illness measures exceeded 0.40, supporting concurrent validity [19,36]. Evidence from analyses based on several different outcomes [33,35] supports predictive validity. Construct validity was studied by comparing scores in two different groups showing expected differences [19,33] and relationships with several other variables showing small but positive associations in the anticipated directions [34,36] (Table 1).

The Duke Severity of Illness (DUSOI) index [37] was developed to assess ambulatory primary care patients, based on patient records, but has also been modified so that it can be used in direct contact between patient and clinician. First, all health problems are identified. For every health problem four domains (symptoms, complications, prognosis without treatment, and treatability) are rated on a five-point scale. The data support concurrent [37], predictive [38], and construct [37] validity. Test-retest [37] and interrater reliability [37–39] are fair, and intrarater reliability is fair to good [37–39] (Table 1).

The Hallstrom index was specifically developed to assist in predicting the outcome of cardiac arrest [40]. It consists of a chronic factor (CF) and a symptom factor (SF). The CF is the number of present conditions from a set of 10 conditions, and the SF is the number of present symptoms from a set of six symptoms related to cardiac disease. A low rank correlation between the two factors was found (0.22,  $P < .001$ ), suggesting that the two scales assess two different concepts. The limited available data provide some support for predictive and construct validity [40] (Table 1).

The Hurwitz index was used in a study that was designed to assess the influence of comorbidity on the type of care (primary, specialist, chiropractic or other) that patients seek for their back problems [41]. Every patient was classified as having either no comorbidity, nondisabling comorbidity, or disabling comorbidity. The index was only able to distinguish between medical and chiropractic care, thus providing limited support for predictive validity [41] (Table 1).

The Index of Coexistent Disease (ICED) [42] consists of two different dimensions, one measuring the disease severity of 14 categories of comorbid diseases (ICED-DS), and

one measuring the “overall functional severity” (disability) caused by comorbidity (ICED-FS). Scores are based on an explicit list of symptoms, signs, and laboratory tests. All information contained in a medical chart can be used to calculate a score. The data support concurrent [4,21,42] and predictive validity [42]. Intrarater reliability is good, and interrater reliability is fair [42,43] (Table 1).

The Incalzi index consists of 52 conditions, each weighted according to its relative risk (RR) for mortality [44]. An Incalzi age index can be computed by adding two, three, or four points to the score of patients aged 76 to 85 years, 86 to 95 years, and over 95 years, respectively [44]. Predictive validity was shown for both the dichotomized (cutoff values were identified on ROC curves) Incalzi and the Incalzi age index for predicting mortality. Showing that the mortality rate for patients with scores above the 75th percentile was higher than for those with scores under the 75th percentile provide further support [44]. According to the authors, interrater reliability was good (data not presented in their article) [44] (Table 1).

The Kaplan index uses two forms of classification, focusing on the type of comorbidity and the pathophysiologic severity of the present comorbid conditions, respectively. The type of comorbidity can be classified as vascular (hypertension, cardiac disorders, peripheral vascular disease, retinopathy, and cerebrovascular disease) or nonvascular (lung, liver, bone, and nondiabetic renal diseases). Pathophysiologic severity is rated on a four-point scale, ranging from 0 (no, or easy to control comorbidity) to 3 (recent full decompensation of comorbid condition). The rating of the most severe condition determines the overall comorbidity score. Scores for vascular and nonvascular comorbidity can be calculated, based on the most severe condition in each subscale. There are two adaptations, the Modified Medical Comorbidity Index (MMCI) and the Adult Comorbidity Evaluation 27 (ACE-27), available [45,46]. Although the adaptations look promising, there are, to our knowledge, no articles published in which a detailed description of either the (content) validity or the reliability is given. The ability of the Kaplan index to predict mortality was studied [5,14,47]. The results support predictive validity (Table 1).

The Liu index consists of 38 conditions, and was specifically constructed for use in stroke outcome research [18]. Every condition is rated on a six-point scale, ranging from 0 (not present) to 5 (active rehabilitation is contraindicated). The Liu index has been compared with the Charlson index, yielding a borderline fair correlation [18]. This result provides some support for concurrent validity, because, given the different objective of the Charlson index, this correlation should neither be too low nor too strong. The Liu index is able to predict scores on the Functional Independence Measure (FIM) and Length of Stay (LOS), supporting predictive validity [18]. Fair correlation coefficients in the anticipated directions between the Liu index and some other variables provide support for construct validity [18,48]. Interrater reliability is good [18] (Table 1).

The Shwarz index consists of 21 weighted conditions, selected from 52 conditions that were derived from the literature on the basis of their positive relationship with mortality or their negative influence on the treatment of the primary condition, using a regression model that was made to predict costs [49]. The Shwarz index can be used with medical records and with databases that use ICD-9-CM codes. Data supporting predictive validity were obtained from regression models predicting costs for the Shwarz index [49], using subgroup analyses and other data sets [49]. Models based on data from medical records performed better than models based on ICD-9-CM codes [49]. According to the authors, intrarater reliability was high (data not presented in their article) [49] (Table 1).

#### 4. Discussion

Measuring comorbidity is an aspect of research that is receiving increasing attention in the literature. Several authors have discussed and compared the use of various selected methods to measure comorbidity [50–53]. This review describes methods that can be used to measure comorbidity in clinical research, without limiting the focus to certain index diseases or diagnostic groups. Thirteen different methods were identified. Six indexes used a carefully developed list of clearly defined diagnoses (BOD, Charlson Index, Hallstrom Index, Incalzi Index, Liu Index, and Shwartz Index). Three indexes rated comorbidity burden by using a system that assessed the effect of comorbid conditions on specific body systems (CIRS, ICED, and Kaplan Index). Two indexes rated comorbidity on a three- or four-point scale using very broad categories (Cornoni-Huntley Index and Hurwitz Index). Two methods used every present condition to calculate a score: one simply counted the number of present comorbid conditions (Disease count) and the other calculated a summary score based on weighted scores for every present comorbid condition (DUSOI).

Although all these methods were developed to measure comorbidity, in the current literature there is no consensus regarding the definition of comorbidity. According to Feinstein [1], comorbidity is defined as “any distinct additional entity that has existed or may occur during the clinical course of a patient who has the index disease under study.” Another definition of comorbidity is the cooccurrence of multiple diseases in one person [51]. In this respect, Van den Akker et al. [54] made a useful distinction between, on the one hand, multimorbidity (i.e., the cooccurrence of multiple chronic or acute diseases and medical conditions in one person) and, on the other hand, comorbidity as defined by Feinstein [1]. By definition, for research on multimorbidity no index disease is used, whereas for comorbidity research an index disease is obligatory. Some of the methods identified can be used for both purposes, depending on whether the focus is on measuring the total burden of diseases in a patient (generic multimorbidity measures) or the burden of comorbid diseases in addition to the condition of

interest (generic comorbidity measures). In the latter case the index disease is omitted from the comorbidity measurement. Other methods, such as the Kaplan, the Liu, the Coronni-Huntley, and the Hallstrom indexes, were specifically developed to measure comorbid diseases in addition to one specific index disease (disease specific comorbidity measures). The other nine methods are generic measures.

Comorbidity indexes first identify the present comorbid diseases and subsequently apply weights or (pathophysiologic) severity ratings for these diseases. The technique of applying weights or (pathophysiologic) severity ratings is very valuable. There is evidence that correcting for comorbidity by simply counting the number of existing diseases leads to another conclusion, than correcting for comorbidity by comorbidity indexes that use weights or (pathophysiologic) severity ratings [55].

Although there is a growing body of evidence that comorbidity, as a disease count or an index, is an independent predictor of several outcomes [33,35,56], relatively little is known about the effect of individual disease combinations on the outcome of interest [50,52,53]. A few authors studied the effect of combinations of individual diseases on disability [56,57]. They showed that the effects of some disease combinations on disability were additive, whereas the effects of other disease combinations were synergistic, leading to more disability than would be expected on the basis of addition. For this type of research, which studies the prognosis of comorbidity and multimorbidity, large numbers of patients are required. Continuing this research to include several index diseases and outcomes could lead to the identification of comorbid conditions that are particularly relevant for one specific index disease and the chosen outcome. These comorbid conditions can subsequently be used to develop disease-specific comorbidity measures.

The development of a comorbidity measure is influenced by the population and outcome used [52]. For the Charlson index it was shown that other weights would have been applied if it had been developed for a different population [8,9,12]. Weights were based on the relative risk of dying, and were used to indicate that not all comorbid conditions have the same impact on the total comorbidity burden. What would have happened if they had chosen another outcome measure? It is likely that the weights would have been very different. Take, for example, osteoarthritis. Weights derived from regression analysis, using mortality as the outcome, will probably be very different from those using mobility as the outcome. These influences should be taken into account when selecting an appropriate comorbidity measure.

Commonly used methods to obtain data that can be used to score comorbidity are interviews, questionnaires, physical examinations, medical chart reviews, and coded databases. The completeness of data obtained from interviews and questionnaires depends on the ability of patients to adequately recall the diseases they suffer from. This ability is strongly influenced by the knowledge and memory of the patient [6]. Although the source of information, i.e., the patient, is the same for interviews and questionnaires, the correlation between scores based on interviews, and questionnaires

ranges from 0.45 to 0.63 [6,52]. One advantage of these two methods is that they are easy to apply in settings in which there is no access to detailed patient records. Medical chart reviews probably yield the most complete data [22,58–60], provided that all charts that exist for one patient are collected. Collecting all the charts and screening the content for relevant data can be rather time consuming, thus increasing the administrative burden. The usefulness of coded databases for the assessment of comorbidity has been the subject of several articles [59,61–63]. A major problem is the limited space available for recording present diseases, and when there are multiple diagnoses, a selection must be made. More serious diseases, the disease for which the patient was admitted, and complications during hospitalization have a higher chance of being recorded than chronic conditions [15,59], introducing substantial bias. Increasing the number of diagnoses might limit this bias, although it is doubtful whether this will solve the problem [59]. Studies comparing scores derived from medical records with those derived from large ICD-9 code-based administrative databases, showed that data derived from medical records are more complete than those derived from administrative databases, especially with regard to asymptomatic diseases [14,22,52,59,62]. Administrative databases yield data for large patient groups, but for smaller studies data from medical records should preferably be used, although data from interviews or questionnaires are a useful alternative.

In conclusion, the Charlson Index, the CIRS, the ICED, and the Kaplan Index are valid and reliable methods to measure comorbidity that can be used in clinical research. For the other indexes, insufficient data on the clinimetric properties are available to assess their validity and reliability. When mortality is the outcome of interest, the Charlson Index has been studied most extensively, and there are several adaptations available. An advantage of the CIRS is the close resemblance to common clinical practice: it is structured according to clinically relevant body systems and uses a clear severity ranking that is clinically sound. Given the good validity and reliability, the CIRS seems a very useful comorbidity measure in clinical research. The ICED is the only measure included in this review that has a two-dimensional structure, measuring both pathophysiologic disease severity and disability. This might be particularly useful in studies in which mortality and disability are the outcomes of interest. The Kaplan index was specifically developed for use in diabetes research, and contains clinically relevant information. It makes a distinction between vascular and nonvascular comorbidity, and uses severity rankings based on parameters derived from common clinical practice. This good face validity, together with the good psychometric properties, makes the Kaplan Index a useful comorbidity index in clinical diabetes research.

## References

- [1] Feinstein AR. The pre-therapeutic classification of co-morbidity in chronic disease. *J Chronic Dis* 1970;23:455–68.

- [2] Streiner DL, Norman GR. Health Measurement Scales. A practical guide to their development and use. Vol. 2. Oxford: Oxford University Press; 1995. p. 1–230.
- [3] Fleiss JL. Reliability of measurement. The design and analysis of clinical experiments. New York: John Wiley & Sons; 1986. p. 1–32.
- [4] Mulrow CD, Gerety MB, Cornell JE, Lawrence VA, Kanten DN. The relationship between disease and function and perceived health in very frail elders. *J Am Geriatr Soc* 1994;42:374–80.
- [5] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
- [6] Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW. Can comorbidity be measured by questionnaire rather than medical record review? *Med Care* 1996;34:73–84.
- [7] Roos LLJ, Nicol JP, Cageorge SM. Using administrative data for longitudinal research: comparisons with primary data collection. *J Chronic Dis* 1987;40:41–9.
- [8] Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* 1992;45:613–9.
- [9] Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *J Clin Epidemiol* 1993;46:1075–9; discussion 1081–9.
- [10] D'Hoore W, Sicotte C, Tilquin C. Risk adjustment in outcome assessment: the Charlson comorbidity index. *Methods Inf Med* 1993;32:382–7.
- [11] D'Hoore W, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson comorbidity index with administrative data bases. *J Clin Epidemiol* 1996;49:1429–33.
- [12] Ghali WA, Hall RE, Rosen AK, Ash AS, Moskowitz MA. Searching for an improved clinical comorbidity index for use with ICD-9-CM administrative data. *J Clin Epidemiol* 1996;49:273–8.
- [13] Cleves MA, Sanchez N, Draheim M. Evaluation of two competing methods for calculating Charlson's comorbidity index when analyzing short-term mortality using administrative data. *J Clin Epidemiol* 1997;50:903–8.
- [14] Newschaffer CJ, Bush TL, Penberthy LE, Bellantoni M, Helzlsouer K, Diener-West M. Does comorbid disease interact with cancer? An epidemiologic analysis of mortality in a cohort of elderly breast cancer patients. *J Gerontol* 1998;53:M372–8.
- [15] Librero J, Peiro S, Ordinanza R. Chronic comorbidity and outcomes of hospital care: length of stay, mortality, and readmission at 30 and 365 days. *J Clin Epidemiol* 1999;52:171–9.
- [16] Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *J Clin Epidemiol* 1994;47:1245–51.
- [17] Melchiorre PJ, Findley T, Boda W. Functional outcome and comorbidity indexes in the rehabilitation of the traumatic versus the vascular unilateral lower limb amputee. *Am J Phys Med Rehabil* 1996;75:9–14.
- [18] Liu M, Domen K, Chino N. Comorbidity measures for stroke outcome research: a preliminary study. *Arch Phys Med Rehabil* 1997;78:166–72.
- [19] Rochon PA, Katz JN, Morrow LA, McGlinchey-Bertho R, Ahlquist MM, Sarkarati M, et al. Comorbid illness is associated with survival and length of hospital stay in patients with chronic disability. A prospective comparison of three comorbidity indices. *Med Care* 1996;34:1093–101.
- [20] Extermann M, Overcash J, Lyman GH, Parr J, Balducci L. Comorbidity and functional status are independent in older cancer patients. *J Clin Oncol* 1998;16:1582–7.
- [21] Gabriel SE, Crowson CS, O'Fallon WM. A comparison of two comorbidity instruments in arthritis. *J Clin Epidemiol* 1999;52:1137–42.
- [22] Kieszak SM, Flanders WD, Kosinski AS, Shipp CC, Karp H. A comparison of the Charlson comorbidity index derived from medical record data and administrative billing data. *J Clin Epidemiol* 1999;52:137–42.
- [23] West DW, Satariano WA, Ragland DR, Hiatt RA. Comorbidity and breast cancer survival: a comparison between black and white women. *Ann Epidemiol* 1996;6:413–9.
- [24] Skiest DJ, Rubinstien E, Carley N, Gioiella L, Lyons R. The importance of comorbidity in HIV-infected patients over 55: a retrospective case-control study. *Am J Med* 1996;101:605–11.
- [25] Poses RM, McClish DK, Smith WR, Bekes C, Scott WE. Prediction of survival of critically ill patients by admission comorbidity. *J Clin Epidemiol* 1996;49:743–7.
- [26] Beddhu S, Bruns FJ, Saul M, Seddon P, Zeidel ML. A simple comorbidity scale predicts clinical outcomes and costs in dialysis patients. *Am J Med* 2000;108:609–13.
- [27] Linn BS, Linn MW, Gurel L. Cumulative illness rating scale. *J Am Geriatr Soc* 1968;16:622–6.
- [28] Miller MD, Paradis CF, Houck PR, Mazumdar S, Stack JA, Rifai AH, et al. Rating chronic medical illness burden in geropsychiatric practice and research: application of the Cumulative Illness Rating Scale. *Psychiatry Res* 1992;41:237–48.
- [29] Conwell Y, Forbes NT, Cox C, Caine ED. Validation of a measure of physical illness burden at autopsy: the Cumulative Illness Rating Scale. *J Am Geriatr Soc* 1993;41:38–41.
- [30] Waldman E, Potter JF. A prospective evaluation of the Cumulative Illness Rating Scale. *Aging* 1992;4:171–8.
- [31] Parmelee PA, Thuras PD, Katz IR, Lawton MP. Validation of the Cumulative Illness Rating Scale in a geriatric residential population. *J Am Geriatr Soc* 1995;43:130–7.
- [32] Cornoni-Huntley JC, Foley DJ, Guralnik JM. Co-morbidity analysis: a strategy for understanding mortality, disability and use of health care facilities of older people. *Int J Epidemiol* 1991;20(Suppl 1):S8–17.
- [33] Page-Shafer K, Delorenze GN, Satariano WA, Winkelstein W Jr. Comorbidity and survival in HIV-infected men in the San Francisco Men's Health Survey. *Ann Epidemiol* 1996;6:420–30.
- [34] Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care* 1998;36:8–27.
- [35] Stineman MG, Escarce JJ, Tassoni CJ, Goin JE, Granger CV, Williams SV. Diagnostic coding and medical rehabilitation length of stay: their relationship. *Arch Phys Med Rehabil* 1998;79:241–8.
- [36] Gross PA, Stein MR, van Antwerpen C, DeMauro PJ, Boscamp JR, Hess W, et al. Comparison of severity of illness indicators in an intensive care unit. *Arch Intern Med* 1991;151:2201–5.
- [37] Parkerson GR Jr, Broadhead WE, Tse CK. The Duke Severity of Illness Checklist (DUSOI) for measurement of severity and comorbidity. *J Clin Epidemiol* 1993;46:379–93.
- [38] Parkerson GR Jr, Broadhead WE, Tse CK. Health status and severity of illness as predictors of outcomes in primary care. *Med Care* 1995;33:53–66.
- [39] Parkerson GR Jr, Bridges-Webb C, Gervas J, Hofmans O, Lamberts H, Froom J, et al. Classification of severity of health problems in family/general practice: an international field trial. *Fam Pract* 1996;13:303–9.
- [40] Hallstrom AP, Cobb LA, Yu BH. Influence of comorbidity on the outcome of patients treated for out-of-hospital ventricular fibrillation. *Circulation* 1996;93:2019–22.
- [41] Hurwitz EL, Morgenstern H. The effects of comorbidity and other factors on medical versus chiropractic care for back problems. *Spine* 1997;22:2254–63.
- [42] Greenfield S, Apolone G, McNeil BJ, Cleary PD. The importance of co-existent disease in the occurrence of postoperative complications and one-year recovery in patients undergoing total hip replacement. Comorbidity and outcomes after hip replacement. *Med Care* 1993;31:141–54.
- [43] Imamura K, Mckinnon M, Middleton R, Black N. Reliability of a comorbidity measure: the Index of Co-Existent Disease (ICED). *J Clin Epidemiol* 1997;50:1011–6.
- [44] Incalzi RA, Capparella O, Gemma A, Landi F, Bruno E, Di Meo F, et al. The interaction between age and comorbidity contributes to predicting the mortality of geriatric patients in the acute-care hospital. *J Intern Med* 1997;242:291–8.



- [45] Piccirillo JF. Impact of comorbidity and symptoms on the prognosis of patients with oral carcinoma. *Arch Otolaryngol Head Neck Surg* 2000;126:1086–8.
- [46] Piccirillo JF. Importance of comorbidity in head and neck cancer. *Laryngoscope* 2000;110:593–602.
- [47] Kaplan MH, Feinstein AR. The importance of classifying initial comorbidity in evaluating the outcome of diabetes mellitus. *J Chronic Dis* 1974;27:387–404.
- [48] Liu M, Tsuji T, Tsujiuchi K, Chino N. Comorbidities in stroke patients as assessed with a newly developed comorbidity scale. *Am J Phys Med Rehabil* 1999;78:416–24.
- [49] Shwartz M, Iezzoni LI, Moskowitz MA, Ash AS, Sawitz E. The importance of comorbidities in explaining differences in patient costs. *Med Care* 1996;34:767–82.
- [50] Gijsen R, Hoeymans N, Schellevis FG, Ruwaard D, Satariano WA, van den Bos GA. Causes and consequences of comorbidity: a review. *J Clin Epidemiol* 2001;54:661–74.
- [51] Guralnik JM. Assessing the impact of comorbidity in the older population. *Ann Epidemiol* 1996;6:376–80.
- [52] Extermann M. Measurement and impact of comorbidity in older cancer patients. *Crit Rev Oncol Hematol* 2000;35:181–200.
- [53] Extermann M. Measuring comorbidity in older cancer patients. *Eur J Cancer* 2000;36:453–71.
- [54] van den Akker M, Buntinx F, Metsemakers JF, Roos S, Knottnerus JA. Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases. *J Clin Epidemiol* 1998;51:367–75.
- [55] Concato J, Horwitz RI, Feinstein AR, Elmore JG, Schiff SF. Problems of comorbidity in mortality after prostatectomy. *JAMA* 1992;267:1077–82.
- [56] Verbrugge LM, Lepkowski JM, Imanaka Y. Comorbidity and its impact on disability. *Milbank Q* 1989;67:450–84.
- [57] Fried LP, Bandeen-Roche K, Kasper JD, Guralnik JM. Association of comorbidity with disability in older women: the Women's Health and Aging Study. *J Clin Epidemiol* 1999;52:27–37.
- [58] Malenka DJ, McLerran D, Roos N, Fisher ES, Wennberg JE. Using administrative data to describe casemix: a comparison with the medical record. *J Clin Epidemiol* 1994;47:1027–32.
- [59] 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;267:2197–203.
- [60] Iezzoni LI, Ash AS, Coffman GA, Moskowitz MA. Predicting in-hospital mortality. A comparison of severity measurement approaches. *Med Care* 1992;30:347–59.
- [61] Wang PS, Walker A, Tsuang M, Orav EJ, Levin R, Avorn J. Strategies for improving comorbidity measures based on Medicare and Medicaid claims data. *J Clin Epidemiol* 2000;53:571–8.
- [62] Humphries KH, Rankin JM, Carere RG, Buller CE, Kiely FM, Spinelli JJ. Co-morbidity data in outcomes research: are clinical data derived from administrative databases a reliable alternative to chart review? *J Clin Epidemiol* 2000;53:343–9.
- [63] Iezzoni LI. The risks of risk adjustment. *JAMA* 1997;278:1600–7.