

# Utilizing national patient-register data to control for comorbidity in prognostic studies

This article was published in the following Dove Press journal:

Clinical Epidemiology

24 October 2014

Number of times this article has been viewed

Yngvar Nilssen<sup>1</sup>  
Trond-Eirik Strand<sup>1</sup>  
Robert Wiik<sup>2</sup>  
Inger Johanne Bakken<sup>3</sup>  
Xue Qin Yu<sup>4,5</sup>  
Dianne L O'Connell<sup>4-7</sup>  
Bjørn Møller<sup>1</sup>

<sup>1</sup>Department of Registration, Cancer Registry of Norway, Oslo, <sup>2</sup>Norwegian Patient Register Department, Norwegian Directorate of Health, Trondheim, <sup>3</sup>Norwegian Institute of Public Health, Oslo, Norway; <sup>4</sup>Cancer Research Division, Cancer Council NSW, <sup>5</sup>School of Public Health, University of Sydney, <sup>6</sup>School of Public Health and Community Medicine, University of New South Wales, Sydney, <sup>7</sup>School of Medicine and Public Health, University of Newcastle, Newcastle, NSW, Australia

**Objective:** To construct an updated comorbidity index (Patient Register Index [PRI]) using national data collections from Norway and compare its predictive ability of 1-year mortality with the Charlson Comorbidity Index (CCI).

**Materials and methods:** Data regarding over 1.11 million patients registered in the Norwegian Patient Register in 2010 and 2011 were used to construct the PRI. The PRI was evaluated by comparing its model fit and discrimination with the CCI.

**Results:** Compared with the CCI, the PRI weights decreased for six, increased for four, and were unchanged for seven diseases. When the PRI was added to the model including age and sex, the age effects were reduced by up to 38% for patients older than 50 years. All measures of model fit improved for the PRI model.

**Conclusion:** Adjustment for comorbidity is especially important for patients 50 years of age or older, and its effect on 1-year mortality is almost comparable to the age effect. The PRI is based on more recent data than the CCI, and is more representative of the general population due to its construction.

**Keywords:** comorbidity, National Patient Register, Charlson Comorbidity Index, predictive ability, patient-register index, prognostic value

## Introduction

Comorbidities are defined as diseases that are concomitant to the disease under study.<sup>1</sup> When survival in medical and epidemiological research is studied, it is important to consider the effect of comorbidity, as it could be a potential confounder or an effect modifier for other prognostic factors.<sup>2-6</sup>

The Charlson Comorbidity Index (CCI) and the Elixhauser method are the most commonly used methods to assess comorbidity.<sup>7,8</sup> The CCI is used more often, is less complex than the Elixhauser method, and was therefore chosen as the reference index in the present study. The CCI was originally developed based on data from all patients admitted to the New York Hospital–Cornell Medical Center during a 1-month period in 1984, with the objective of predicting 1-year mortality. The risk of death associated with each of the 19 predefined diseases included in the CCI was expressed as weights with values of 1, 2, 3, or 6. Summing the weights for all contributing diseases gives the CCI score for each patient. The CCI was validated using an external cohort consisting of 685 breast cancer patients receiving their first treatment in 1962–1969 at Yale New Haven Hospital.<sup>7</sup>

The original CCI has been modified and evaluated by many authors.<sup>9-13</sup> Quan et al developed coding algorithms for constructing the CCI based on the codes of the

Correspondence: Yngvar Nilssen  
Department of Registration, Cancer Registry of Norway, PO Box 5313, Majorstuen, N-0304 Oslo, Norway N-0304, Norway  
Tel +47 23 333 915  
Email yngvar.nilssen@krefregisteret.no

Tenth Revision of the International Classification of Diseases (ICD-10).<sup>14</sup> Sundararajan et al assessed that the version of the CCI created by Quan et al outperformed all the other versions they considered, though all of the ICD-10 versions of the Charlson algorithm performed satisfactorily.<sup>15</sup> However, a persistent challenge for researchers who wish to use the CCI is that the data set of interest may not contain sufficient medical information to assess all the diseases included in the index. One source of such information in many countries is national patient registers. These are based on hospital administrative databases, from which data are often readily available.

Due to better treatment and technological improvements, a patient's risk of death has reduced since the CCI was established in the 1980s, and the index has been updated accordingly. Bottle and Aylin<sup>12</sup> and Quan et al both updated the CCI.<sup>11</sup> Bottle and Aylin used national data from the UK and studied in-hospital mortality, while Quan et al used regional data from the Calgary Health Region of Canada to study 1-year case fatality. However, to our knowledge, no previous study has combined the use of national patient-register data and death within 1 year as the study end point.

The aim of the present study was to evaluate the importance of including a comorbidity measure into analyses when predicting 1-year mortality using data from the Norwegian Patient Register (NPR). A second aim was to construct a modified version of the CCI that can be used with patient-register data (the Patient Register Index, PRI), and compare the predictive ability of the CCI and the PRI, as well as the confounding effect of each index on age. An additional aim was to describe the pattern of diseases included in the CCI by age and sex, based on data from all hospitals in Norway registered in the NPR.

## Materials and methods

### The NPR

The NPR is a national health register covering all sectors of the specialized health care services. Reporting to NPR is mandatory, and the register includes data on all patients treated in Norwegian government-funded hospitals. Personal identification numbers have been reported to the NPR from 2008 onwards. This enables researchers and policy makers to follow the disease trajectory of patients between sectors and hospitals. Also, alignment of data and validation with other national health registries is made feasible.

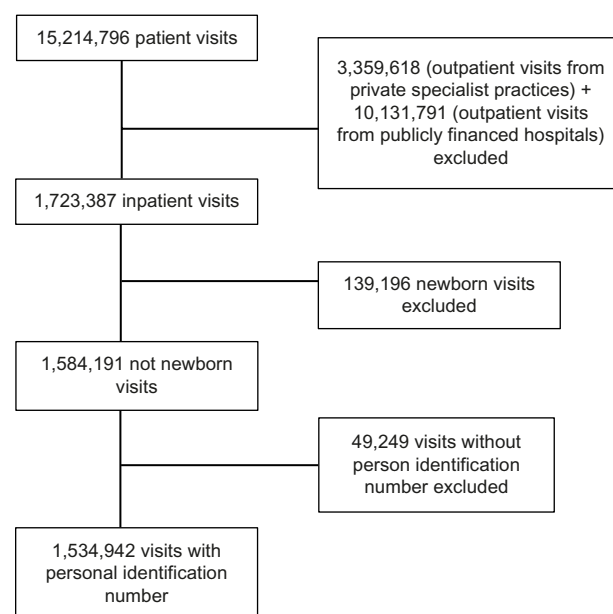
The study sample was selected from the NPR data reported from all hospitals, which consisted of three main data sources for statistics. The first source was visits for medical treatment

for in- and outpatients at publicly financed hospitals. The two other sources were private hospitals and private specialist practices. The government purchases medical treatment from private hospitals and specialists practices as a supplement to services at the public hospitals. The NPR does not include data on privately financed hospital treatments.

The basic data unit in the NPR is hospital visits. However, when a patient is transferred between wards at the same hospital, the record is aggregated. Each episode of national hospital data contains one or more diagnoses, coded according to the ICD-10.

### Study sample

To provide information on the total disease history of the patient, the personal identification number was used to link episodes of treatment registered at different hospitals, sectors, and years. Initially, all patients in the NPR registered with a hospital visit in 2010 or 2011 were identified, constituting 15,214,796 visits. Figure 1 shows an overview of exclusions of patient visits in the study. Three subgroups of the initial data suffered from incomplete reporting of personal identification numbers. The first group was outpatients at private specialist practices in 2010 and 2011 (3,359,618 visits). The reason for incomplete reporting in this group was mainly related to technical limitations in their information-technology systems. To avoid any selection bias, we then also excluded 10,131,791 outpatient visits registered at publicly financed hospitals. However, we decided to keep



**Figure 1** Flowchart shows the included and excluded patients in the study population.

outpatient visits with surgical procedure codes, because this activity mainly is performed at publicly financed hospitals or publicly financed private hospitals. The second group of visits excluded due to incomplete reporting were data on newborn babies (139,196 visits), since a large proportion had not yet received a personal identification number at the time of registration. The third group was any other patient visits that were lacking a correctly reported personal identification number (49,249 visits). After these exclusions, the data were aggregated, resulting in 1,534,942 eligible visits, corresponding to 1,113,341 unique patients.

The first hospital visit during the study period was defined as the patient's index visit. The end of follow-up was defined as the date of death if the patient died within 1 year of the index visit, or the date 1 year after the index visit (censoring date). Sex, age category, and ICD-10 codes (truncated to three digits for the index visit) were obtained from the NPR. To avoid the possibility of tracing back to individual patients, we received age information in 5-year age-group categories. Duration in days from the index visit to end of follow-up, vital status, and history of the diseases (dichotomous variables) included in the CCI that were listed up to 1 year prior to the index visit were obtained from the NPR. The NPR routinely obtains information on vital status and date of death from the National Population Register.<sup>16</sup> The data file obtained from the NPR was anonymized, and thus no requirement for regulatory ethical approval in Norway was needed.

The original CCI included 19 diseases, but for the purposes of the present study, leukemia and lymphoma were included in the disease category "any malignancy", as done by other authors.<sup>9–11</sup> These 17 CCI diseases were identified by the ICD-10 coding defined by Quan et al in 2005.<sup>14</sup> A primary diagnostic code is the code for the main medical condition causing the admission. Secondary diagnostic codes are codes for diseases that exist at the same time as the primary disease, or diseases that develop and need examination or treatment during the admission. Information from both primary and secondary diagnostic codes was obtained for each patient.

## Statistical analysis

The PRI weights were estimated using a Cox regression model of time from index visit to death with follow-up censored at 1 year, adjusted for sex, 5-year age categories (0–4 years ... 95+ years) and 17 CCI disease indicators. An algebraically correct method was used to define new weights for the CCI diseases by summing the regression

coefficients, not the hazard ratios (HRs).<sup>17</sup> The PRI was constructed by multiplying the regression coefficients by a scaling constant –  $k$ , then rounding it to the nearest integer, and finally summing it over all the CCI diseases. The scaling constant  $k$  was chosen such that the maximum weight for a specific disease in the PRI became the same integer as the maximum weight in the CCI, and thus  $k$  was set to 2.3.

Both the construction and comparison of the PRI and the CCI were done through an internal tenfold cross-validation procedure.<sup>18</sup> Ten subsets of the data were randomly defined, and nine of them were used to develop the PRI, with the last subset used to run the comparison. This was done for all ten subsets, and the results are given as the mean values from the ten runs. To predict 1-year mortality, three logistic regression models were fitted, one including sex and age (base model), one including sex, age, and the CCI (CCI model), and one model including sex, age, and the PRI (PRI model).

Model fit was compared using several measures of global fit. The Bayesian information criterion (BIC) is a likelihood-based measure, and for model fit, the lower the BIC-value, the better the fit.<sup>19</sup> Likelihood ratio chi-square statistics were also used to test model fit. McFadden's  $R^2$  is a measure of improvement in fit over the intercept model.<sup>20</sup> The Brier score was used to directly compare the observed outcomes with the predicted probabilities.<sup>19</sup> The  $C$ -statistic, which is a summary of a model's ability to discriminate between those who do and those who do not experience the outcome, was calculated for each model.<sup>21</sup> This measure is the most commonly used in the medical and epidemiological literature when comparing different comorbidity indices.<sup>11,13,15,22,23</sup> The  $C$ -statistic varies from 0.5, which indicates that the discrimination is due to chance alone, to 1, which indicates perfect discrimination. The general classification of discrimination is "acceptable" ( $c \in [0.7, 0.8]$ ), "excellent" ( $c \in [0.8, 0.9]$ ), and "outstanding" ( $c \in [0.9, 1]$ ).<sup>24</sup> To assess which index changed the predicted value most when added to the base model, movement from the estimated probabilities of the outcome were compared. This was done using the net reclassification improvement (NRI), which is defined as the difference in proportions moving up or down in risk among patients who died and those who survived. We also used the integrated discrimination improvement (IDI), which measures the difference in the mean predicted probabilities between those who died and those who did not in the CCI and PRI models compared to the base model.<sup>25,26</sup> We also calculated how much the probabilities changed on average when a comorbidity index was included in the model:

$$\Delta\hat{p} = \frac{1}{N} \sum_{i=1}^N | \hat{p}_{\text{age,sex},i} - \hat{p}_{\text{age,sex,comorbidity},i} | \quad (1)$$

This quantifies the effect of introducing a comorbidity index into a model. To compare the importance of including age, sex, and a comorbidity index in the model,  $\Delta\hat{p}$  was calculated with and without each of the variables in a subanalysis with patients aged over 50 years who were not admitted to hospital due to any of the 17 CCI diseases at the index visit.

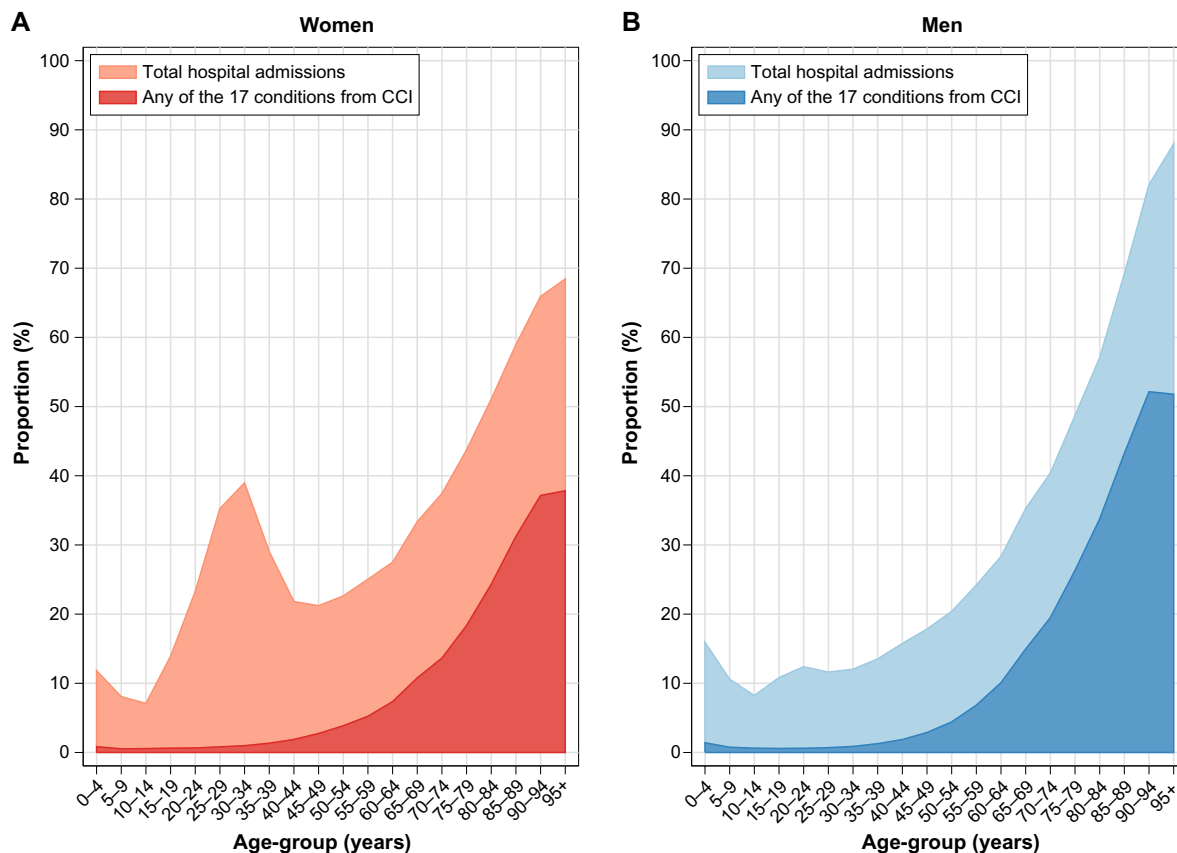
In the description of the disease pattern, the patient group was age-standardized according to the Norwegian standard population in 2011.<sup>27</sup> All analyses were done using Stata 13.<sup>28</sup>

### Results

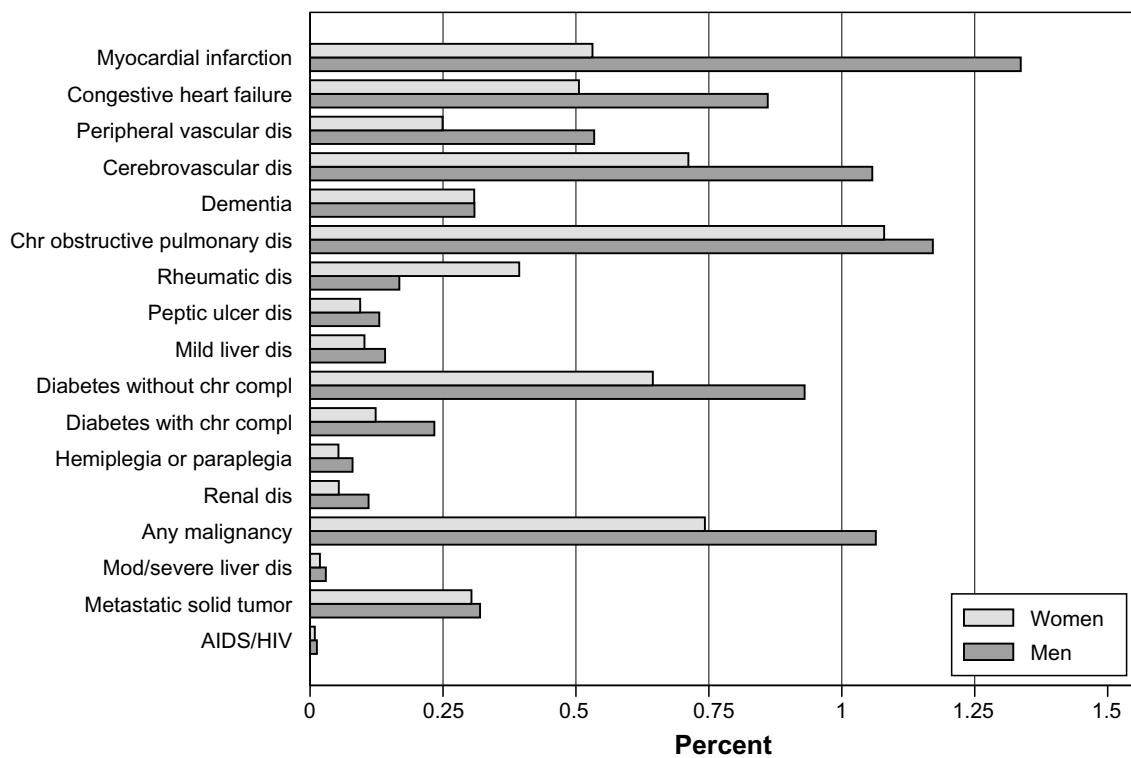
The majority of the 1,113,341 patients in the study were women (57.0%). For 68.1% of the patients, the index visit was a hospitalization. Figure 2 presents the proportion of the Norwegian population with one or more of the CCI diseases in 2010–2011 and the total proportion admitted to Norwegian hospitals in that same period. Twenty-two

percent of the patients were registered with at least one CCI disease, but among men and women below 50 years of age, the proportion was less than 3%. For patients 50 years of age or older, the proportion with CCI diseases increased with age, and the increase was more prominent among men. For both sexes, there was a strikingly high proportion of hospital admission in the youngest age-group (0–4 years). This was mostly due to diseases of the respiratory system (data not shown). For women, there was a peak around 30–34 years (Figure 2) as a result of birth-related admissions (data not shown). There was a small peak for men aged 20–24 years as well, which was mostly attributed to ICD-10 codes within S00-T98 (“Injury, poisoning, and certain other consequences of external causes”, data not shown).

The age-standardized proportions of the Norwegian population with each of the CCI diseases are shown in Figure 3. The five most common CCI diseases for men and women were myocardial infarction, cerebrovascular disease, chronic pulmonary disease, diabetes without chronic complications, and malignancies, but the order of importance differed.



**Figure 2** Total proportion of the Norwegian population admitted to hospitals in 2010–2011 in Norway or with a condition included in the Charlson Comorbidity Index (CCI) present within the previous year, by sex and age-group. **Notes:** Data shown are for (A) women and (B) men.



**Figure 3** Age-standardized proportion (%) of the Norwegian population admitted to hospitals in 2010–2011 in Norway with any of the 17 conditions in the Charlson Comorbidity Index present within the last year, by sex.

**Abbreviations:** dis, disease; chr, chronic; compl, complications; mod, moderate; AIDS, acquired immunodeficiency syndrome; HIV, human immunodeficiency virus.

The overall proportion of deceased patients within 1 year was 4.8% (52,938 patients), and the 1-year risk of death was significantly higher for patients with at least one CCI disease compared to those with none (HR 4.5, confidence interval 4.4–4.6) (Table 1). The proportion of patients with each CCI disease ranged between 0.05% (acquired immunodeficiency syndrome [AIDS]/human immunodeficiency virus [HIV]) and 4.9% (chronic obstructive pulmonary disease [COPD]). Metastatic tumor was associated with the highest hazard, while rheumatic diseases and diabetes without chronic complications were associated with the lowest hazard for 1-year mortality. Across all CCI diseases, the mean CCI score was 0.4, the median number of diseases present in the patients was 0, and the range was 0–8.

When the weights for the CCI diseases in the PRI were compared with those in the CCI, the weights for seven diseases remained unchanged, four increased in magnitude, and six decreased in magnitude. Except for two CCI diseases – moderate-to-severe liver disease and AIDS/HIV – the weights in the PRI only deviated by one if at all compared to those in the CCI. The maximum observed CCI and PRI values among the patients were 12 and 15, respectively. The overall mean CCI was approximately equal to the mean PRI (0.43 versus 0.42).

Table 2 presents the measures of model fit obtained from the internal cross-validation for the three prediction models. The *C*-statistic was marginally better for the PRI model than for the CCI model (91.5 versus 91.3). All measures showed improvement in fit when the PRI was included in the model instead of the CCI (Table 2).

Among patients aged 50 years or older who were not admitted to hospital due to any of the 17 CCI diseases, the ten most common categories for admission are listed in Table 3. ICD-10 code H25, “Senile cataract” (4.2%), and R07, “Pain in throat and chest” (3.0%), were the two most common conditions. The rest contributed with less than 3% each. Figure 4 shows the change in the effect of age when either of the two comorbidity indices was added to the base model for these patients. Including the PRI in the model led to a reduction in the effect of age for all age-groups, ranging from 3.5% in the 55–59-year group to 38% in the 95+ year group. This reduction was greater than the reduction observed for the CCI model.

Using backward elimination from the PRI model for patients 50 years of age or older with an index visit that was not due to a CCI disease, the change in the estimated probability ( $\Delta\hat{p}$ ) was 2.8%, 4.3%, and 0.7% when the PRI, age, and sex were removed, respectively. Similarly, for the CCI

**Table 1** Number and proportion of patients (present and dead) with Charlson Comorbidity Index (CCI) diseases (not mutually exclusive) registered in the Norwegian Patient Register between January 1, 2010 and December 31, 2011

Comorbidity	Number of patients	Proportion with comorbidity (%)	Number dead	Proportion dead (%)	HR	95% CI	CCI	PRI
No	861,553	77.4	12,296	1.4	1.0	–	0	0
Yes	251,788	22.6	40,642	16.1	4.46	4.36–4.55	>0	>0
Total	1,113,341	100.0	52,938	4.8				
<b>Type of comorbidity</b>								
No chr dis	861,553	77.4	12,296	1.4	1.0	–	0	0
Myocardial infarction	44,034	4.0	7,741	17.6	1.25	1.22–1.28	1	1
Congestive heart failure	32,408	2.9	9,745	30.1	2.07	2.02–2.12	1	2
Peripheral vascular dis	18,459	1.7	3,253	17.6	1.36	1.31–1.41	1	1
Cerebrovascular dis	42,425	3.8	7,446	17.6	1.56	1.52–1.60	1	1
Dementia	15,225	1.4	5,391	35.4	2.34	2.27–2.41	1	2
Chronic pulmonary dis	54,410	4.9	7,769	14.3	1.65	1.60–1.69	1	1
Rheumatic dis	14,110	1.3	1,261	8.9	1.03	0.97–1.09	1	0
Peptic ulcer dis	5,408	0.5	931	17.2	1.45	1.36–1.55	1	1
Mild liver dis	6,007	0.5	496	8.3	2.72	2.49–2.98	1	2
Diabetes without chr compl	38,237	3.4	4,654	12.2	1.20	1.17–1.24	1	0
Diabetes with chr compl	8,558	0.8	1,421	16.6	1.44	1.36–1.52	2	1
Hemiplegia or paraplegia	3,265	0.3	224	6.9	1.60	1.40–1.83	2	1
Renal dis	3,905	0.4	536	13.7	1.67	1.54–1.82	2	1
Any malignancy	42,667	3.8	7,564	17.7	2.94	2.86–3.01	2	2
Mod or severe liver dis	1,184	0.1	390	32.9	7.47	6.67–8.27	3	5
Metastatic solid tumor	14,917	1.3	7,777	52.1	13.80	13.46–14.15	6	6
AIDS/HIV	536	0.05	31	5.8	3.61	2.53–5.14	6	3

**Notes:** Hazard ratios (HRs) for 1-year mortality, and 95% confidence intervals (CIs), adjusted for age and sex, together with weights for both the CCI and Patient Register Index (PRI), are shown.

**Abbreviations:** AIDS, acquired immunodeficiency syndrome; HIV, human immunodeficiency virus; dis, disease; chr, chronic; compl, complications; mod, moderate.

model,  $\hat{\Delta}p$  was 2.6%, 4.7%, and 0.7% when the CCI, age, and sex were removed, respectively.

## Discussion

We observed that for patients 50 years of age or older, comorbidity based on NPR data was almost equally important as age for predicting 1-year mortality. Measured by the proportion of the Norwegian population with a CCI disease registered in the NPR, the disease burden increased with age, reaching 52% and 38% in the oldest age-group for men and women, respectively. The weights for four CCI diseases were higher in the PRI compared to the corresponding weights in the CCI, while the weights for six diseases were lower in the PRI than in the CCI.

The PRI was marginally better than the CCI in predicting 1-year mortality. The *C*-statistics showed that the performance of both indices was rated as “outstanding” in discriminating, according to the standard *C*-statistic classification, between those who died and those who did not.<sup>24</sup> The base model produced a *C*-statistic of 0.869, while the *C*-statistic increased to 0.913 and 0.915 when the CCI and the PRI were included in the model, respectively. Therefore, the majority of the ability to discriminate came from sex and age, but both of the indices increased the discrimination ability further, which is

in agreement with the findings of Gabbe et al<sup>29</sup> and Kilgore et al.<sup>30</sup> For patients over 50 years of age, we observed that the PRI was almost as important as age, while the CCI was relatively less important than the PRI. In addition, a reduction in the effect of age for those over 50 years of age increased with age toward 38% when the PRI was added to the base model, showing the importance of including information regarding comorbidity when predicting the probability of death, especially in elderly patients.

The lack of data from privately financed hospitals could cause concern of possible introduced bias. However, in 2008 the total use of privately financed hospitals in Norway was approximately 0.5% of the total health care service.<sup>31,32</sup> Therefore, excluding these patients would probably not have induced any major bias. Further, we only included inpatients and those outpatients with a surgical procedure code. The exclusion of all the outpatient visits with only medical procedures or no procedures at all probably led to an underestimation of the prevalence of the least severe diseases (those who did not need a hospitalization). The main aim in this study, however, was not to estimate the prevalence of a disease, but to construct a valid modification of the CCI. The private specialist practices are not situated evenly throughout the country. If we had included all outpatient

**Table 2** Global measures of model fit, discrimination, and calibration statistics

Model evaluation measures	Base	CCI model	PRI model
<b>Global measures</b>			
LR $\chi^2$	10,556	14,623	14,834
P-value*	<0.01	<0.01	<0.01
BIC	32,269	28,213	28,003
R <sup>2</sup> (%)	24.8%	34.3%	34.8%
Brier (%)	3.9%	3.5%	3.5%
<b>Discrimination</b>			
C-index (%)	86.9%	91.3%	91.5%
NRI (%)	–	74.5%	84.0%
IDI (%)	–	9.9%	10.7%
$\Delta\hat{p}^\dagger$	–	2.6%	2.8%

**Notes:** Models include the base model with sex and age, the model including sex, age, and Charlson Comorbidity Index (CCI), and the model including sex, age, and the Patient Register Index (PRI), using an internal tenfold cross-validation procedure.

\*P-value for LR test between the base model and a model including CCI or PRI;  
 $\dagger \Delta\hat{p} = \frac{1}{N} \sum_{i=1}^N |\hat{p}_{\text{Page, sex, i}} - \hat{p}_{\text{Page, sex, comorbidity, i}}|$

**Abbreviations:** LR, likelihood ratio; BIC, Bayesian information criterion; NRI, net reclassification improvement; IDI, integrated discrimination improvement.

visits, this would have introduced a possible regional bias using the PRI.

The Norwegian Prescription Database reported that approximately 156,000 (3.1%) of Norwegians used medication for diabetes in 2011. The total prevalence of diabetes in Norway is uncertain. In addition to those using medications, there is an unknown number of patients with type 2 diabetes who are managed through lifestyle changes only. However, a Norwegian study reported that 4.3% of the population suffered from diabetes in the period 2006–2008.<sup>33</sup> In our data, we identified approximately 47,000 cases of diabetes treated in hospitals during the study period (1% of the Norwegian population). The Norwegian Institute of Public Health has reported that the prevalence of COPD in Norway is around 200,000 (4%).<sup>34</sup> In our study, 54,410 patients admitted to hospital in 2010–2011 were either treated for COPD or had it registered in their disease history. Lastly, the Cancer Registry of Norway reported a cancer prevalence of 215,000 in 2011.<sup>35</sup> The observed prevalence from the NPR is measured using a combination of prevalence and incidence, since it includes patients with either a newly diagnosed cancer or a diagnosis in the past in continuous need of treatment. However, considering any malignancy and metastatic tumor as one group in the NPR data, there were 57,584 cancer patients. Many patients with COPD and cancer do not need an annual or biannual consultation at a hospital, either because they are well controlled or considered cured from their disease, or since their assigned general practitioner handles regular

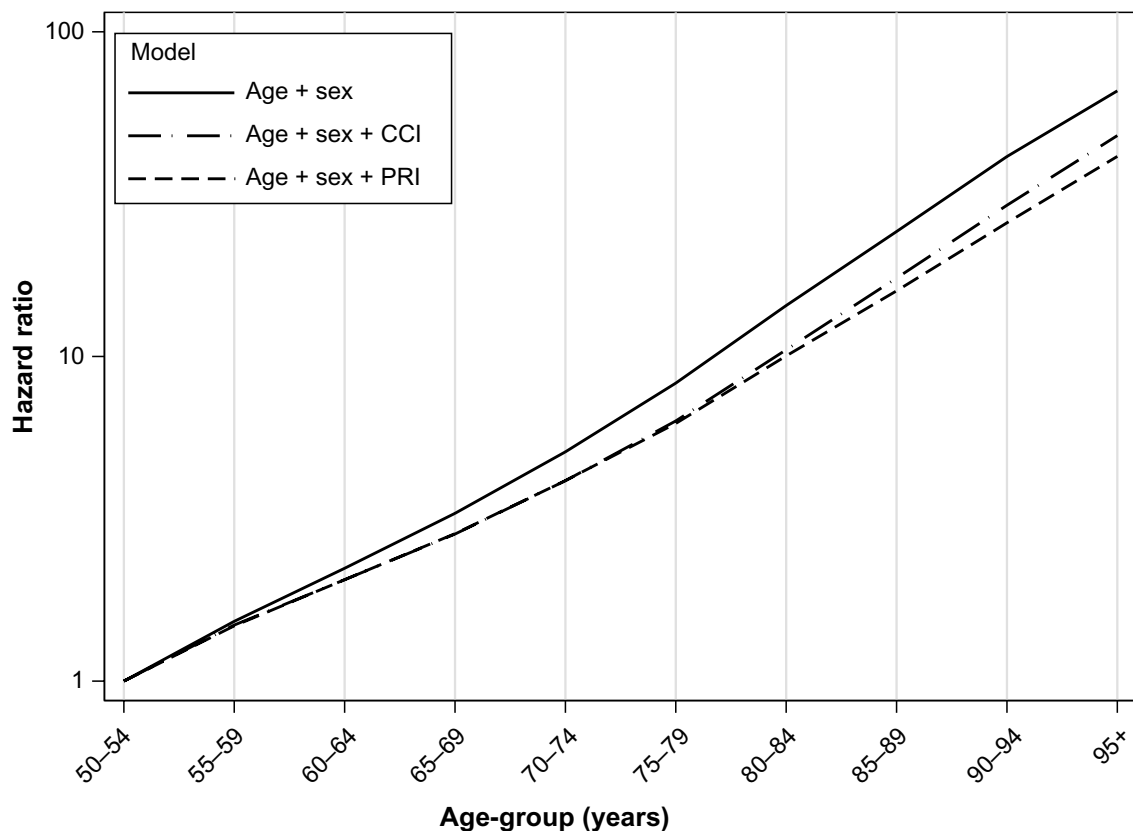
**Table 3** Number and proportion of patients aged 50 years or older not admitted to hospital for any of the 17 Charlson Comorbidity Index diseases

ICD-10	Classification	Number	Proportion (%)
M00-M99	Diseases of the musculoskeletal system and connective tissue	75,156	17.0
S00-T98	Injury, poisoning, and certain other consequences of external causes	57,379	13.0
I00-I99	Diseases of the circulatory system	46,757	10.6
R00-R99	Symptoms, signs, and abnormal clinical and laboratory findings not elsewhere classified	39,821	9.0
K00-K93	Diseases of the digestive system	37,648	8.5
H00-H59	Diseases of the eye and adnexa	35,662	8.1
N00-N99	Diseases of the genitourinary system	35,563	8.0
J00-J99	Diseases of the respiratory system	26,589	6.0
C00-D48	Neoplasms	20,293	4.6
G00-G99	Diseases of the nervous system	19,138	4.3
	Other	48,742	11.0
	Total	442,748	100.0

**Abbreviation:** ICD-10, Tenth Revision of the International Classification of Diseases.

issues, explaining why the numbers in the NPR from 2010 to 2011 are lower than the corresponding published prevalence estimates.<sup>33,34</sup> Bakken et al showed good agreement between the data included in the NPR and the Cancer Registry of Norway.<sup>36</sup> It is reasonable to believe that the patients we identified in the NPR had more severe disease and thus poorer prognosis than those suffering from the same diseases, but not admitted to hospital.

The uniqueness of this study is that we used two alternative methods (NRI and IDI) for comparing the quality of the reclassification of patients. None of the previous papers describing modifications of a comorbidity index has to our knowledge used these predictive probability measures. The NRI was 74.5% and 84.0% for the CCI and PRI, respectively, showing that inclusion of a comorbidity index increased the credibility of a model's predictive ability, and the model including the PRI correctly determined the risk of a larger proportion of patients. Also, here the model including the PRI yielded a higher IDI of 10.7% compared to 9.9% for the model that included the CCI. Therefore, the PRI model has a better ability to reclassify patients than the CCI model. In addition, our defined  $\Delta\hat{p}$  illustrated the importance of including a comorbidity



**Figure 4** The age effects (hazard ratios) relative to the age-group 50–54 years (log scale), for different models based on patients not admitted for any of the 17 Charlson Comorbidity Index (CCI) diseases registered in the Norwegian Patient Register in 2010–2011.

**Abbreviation:** PRI, Patient Register Index.

index in a model, as opposed to using the base model, in predicting the 1-year mortality for patients 50 years of age or older.

The present study is closely related to the work done by Quan et al and Bottle and Aylin.<sup>11,12</sup> Quan et al developed an updated version of the CCI based on the population of Calgary, Canada, and validated it using data from six different nations. A major advantage of the study of Quan et al was that they validated their updated index externally, in addition to being population-based. The advantage of Bottle and Aylin's study was that they developed new empirical weights based on English administrative data in 2007–2008 (with over 5 million records). However, they studied in-hospital mortality only, as individual follow-up was not possible. Our proposed PRI is based on much larger patient numbers than Quan et al's, and more complete follow-up after discharge from hospital compared with that of Bottle and Aylin, which is very important, especially considering that only 35% of the deceased in Norway in 2011 died in hospitals.<sup>37</sup>

When constructing the PRI, diagnoses of the CCI diseases, recorded both in primary and secondary fields from the index visit as well as the history, and recorded up to

1 year prior to the index visit were included. The rationale behind this was that such an index should be based on the most recent disease history available. Pine et al and Ghali et al both showed improved discrimination when conditions present upon admission were included.<sup>38,39</sup> In all subanalyses for patients 50 years of age or older, we excluded data for the patients who were admitted to hospital due to any of the 17 CCI diseases. The rationale behind this was that patients under 50 years of age have a low presence of comorbidities. In addition, we wanted to study comorbidity – diseases in addition to the one causing the index visit – and not morbidity. This study design is similar to the situation researchers meet when they want to add comorbidity information to their study cohort.

One limitation of our study was that we were not able to validate the PRI using external data. Another limitation regards the inclusion criteria of patients: we excluded outpatients registered with only medical procedures or no procedures at all. This excluded many patients and limits the generalizability of the results, but it was necessary due to the large variation regarding the degree of reported personal identification numbers from private specialist practices. Outpatient

consultations are usually scheduled for patients with chronic and non-life-threatening diseases, such as dementia, rheumatism, and diabetes; hence, there is a considerable risk of underestimating the number of patients suffering from these diseases in our study. On the other hand, the strengths of the study are that in contrast to the CCI, the PRI was developed using data from a national register. Secondly, the algebraically correct method of the CCI was used when developing the PRI.<sup>17</sup> Finally, the large sample size and access to complete mortality data are obvious strengths.

## Conclusion

It is of high importance to include a comorbidity index in observational studies of disease prognosis, especially in elderly patients. Both the CCI and the PRI showed a high degree of discrimination, indicating that both have good predictive ability. However, the PRI explained a larger proportion of the observed effect of age, and the PRI weights reflect the patterns in the data from the NPR. We have shown that weighting the CCI for a specific population slightly improves the performance of the CCI. The PRI is by its construction more representative of the general population, and can be generalized to other countries in situations where data from a national patient register are used.

## Disclosure

The authors report no conflicts of interest in this work.

## References

- Feinstein AR. The pre-therapeutic classification of comorbidity in chronic disease. *J Chronic Dis*. 1970;23(7):455–468.
- Birim O, Maat AP, Kappetein AP, van Meerbeeck JP, Damhuis RA, Bogers AJ. Validation of the Charlson comorbidity index in patients with operated primary non-small cell lung cancer. *Eur J Cardiothorac Surg*. 2003;23(1):30–34.
- Piccirillo JF, Tierney RM, Costas I, Grove L, Spitznagel EL Jr. Prognostic importance of comorbidity in a hospital-based cancer registry. *JAMA*. 2004;291(20):2441–2447.
- Lund L, Jacobsen J, Nørgaard M, et al. The prognostic impact of comorbidities on renal cancer, 1995 to 2006: a Danish population based study. *J Urol*. 2009;182(1):35–40; discussion 40.
- Iversen LH, Nørgaard M, Jacobsen J, Laurberg S, Sørensen HT. The impact of comorbidity on survival of Danish colorectal cancer patients from 1995 to 2006 – a population-based cohort study. *Dis Colon Rectum*. 2009;52(1):71–78.
- Degroot V, Beckerman H, Lankhorst G, Bouter L. How to measure comorbidity: a critical review of available methods. *J Clin Epidemiol*. 2003;56(3):221–229.
- 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(5):373–383.
- Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8–27.
- Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol*. 1992;45(6):613–619.
- 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(10):1075–1079; discussion 1081–1090.
- Quan H, Li B, Couris CM, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol*. 2011;173(6):676–682.
- Bottle A, Aylin P. Comorbidity scores for administrative data benefited from adaptation to local coding and diagnostic practices. *J Clin Epidemiol*. 2011;64(12):1426–1433.
- Sarfati D, Gurney J, Stanley J, et al. Cancer-specific administrative data-based comorbidity indices provided valid alternative to Charlson and National Cancer Institute Indices. *J Clin Epidemiol*. 2014;67(5):586–595.
- Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43(11):1130–1139.
- Sundararajan V, Quan H, Halfon P, et al. Cross-national comparative performance of three versions of the ICD-10 Charlson index. *Med Care*. 2007;45(12):1210–1215.
- National Registry [homepage on the Internet]. Available from: [http://www.skatteetaten.no/en/Person/National-Registry/?utm\\_source=1&utm\\_medium=flyttemelding&utm\\_campaign=folkeregisteret](http://www.skatteetaten.no/en/Person/National-Registry/?utm_source=1&utm_medium=flyttemelding&utm_campaign=folkeregisteret). Accessed September 6, 2014.
- Harrell F. Regression coefficients and scoring rules. *J Clin Epidemiol*. 1996;49(7):819.
- Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer; 2009.
- Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer; 2001.
- McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, editor. *Frontiers in Econometrics (Economic Theory and Mathematical Economics)*. Amsterdam: Academic; 1974: 105–142.
- Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007;115(7):928–935.
- Quail JM, Lix LM, Osman BA, Teare GF. Comparing comorbidity measures for predicting mortality and hospitalization in three population-based cohorts. *BMC Health Serv Res*. 2011;11:146.
- Schneeweiss S, Wang PS, Avorn J, Glynn RJ. Improved comorbidity adjustment for predicting mortality in Medicare populations. *Health Serv Res*. 2003;38(4):1103–1120.
- Hosmer DW, Lemeshow S. *Applied Logistic Regression*. Hoboken (NJ): Wiley & Sons; 2000.
- Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27(2):157–172; discussion 207–112.
- Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30(1):11–21.
- Statistics Norway. Statistikkbanken. 2013. Available from: <http://www.ssb.no/statistikkbanken>. Accessed February 10, 2014.
- StataCorp. Stata Statistical Software: Release 13. College Station, TX: StataCorp; 2013.
- Gabbe BJ, Harrison JE, Lyons RA, Edwards ER, Cameron PA. Comparison of measures of comorbidity for predicting disability 12-months post-injury. *BMC Health Serv Res*. 2013;13:30.
- Kilgore ML, Smith W, Curtis JR, et al. Evaluating comorbidity scores based on health service expenditures. *Medicare Medicaid Res Rev*. 2012;2:E1–E8.
- SINTEF. *Omfang og Utvikling av Det Selv-betalende Markedet for Private Spesialisthelsetjenester i Norge*. Trondheim, Norway: SINTEF; 2010. Available from: <http://www.sintef.no/upload/Helse/Rapport%20A15248%20Omfang%20og%20utvikling%20av%20det%20selv-betalende%20markedet.pdf>. Accessed August 7, 2014.

32. Norwegian Directory of Health. *Samdata Spesialisthelsetjenesten 2012*. Oslo: Norwegian Directory of Health; 2013. Available from: <http://helsedirektoratet.no/publikasjoner/samdata-2012/Sider/default.aspx>. Accessed August 7, 2014.
33. Norwegian University of Science and Technology. *Public Health Development: The HUNT Study*. Levanger, Norway: HUNT Forskningscenter; 2011. Available from: <http://www.ntnu.no/documents/10304/1130562/folkehelse-i-ending-huntrapport-2011.pdf>. Accessed February 14, 2014.
34. Norwegian Institute of Public Health. [COPD: facts about chronic obstructive pulmonary disease]. 2013. Available from: <http://www.fhi.no/artikler/?id=55604>. Accessed February 14, 2014. Norwegian.
35. Cancer Registry of Norway. *Cancer in Norway 2011: Cancer Incidence, Mortality, Survival and Prevalence in Norway*. Oslo: Cancer Registry of Norway; 2012.
36. Bakken IJ, Gystad SO, Christensen OO, et al. Comparison of data from the Norwegian Patient Register and the Cancer Registry of Norway. *Tidsskrift Nor Laegeforen*. 2012;132(11):1336–1340.
37. Norwegian Directorate of Health. [Activity data for the specialist health-services 2011]. Oslo: Norwegian Directorate of Health; 2011. Available from: <http://www.helsedirektoratet.no/publikasjoner/aktivitetsdata-for-somatisk-spesialisthelsetjeneste-2011/Publikasjoner/rapport-somatikk-2011.pdf>. Accessed February 14, 2014. Norwegian.
38. Pine M, Jordan HS, Elixhauser A, et al. Enhancement of claims data to improve risk adjustment of hospital mortality. *JAMA*. 2007;297(1):71–76.
39. Ghali WA, Quan H, Brant R. Risk adjustment using administrative data: impact of a diagnosis-type indicator. *J Gen Intern Med*. 2001;16(8):519–524.

## Clinical Epidemiology

### Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

Submit your manuscript here: <http://www.dovepress.com/clinical-epidemiology-journal>

systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Dovepress