Aging Clinical and Experimental Research (2025) 37:172
https://doi.org/10.1007/s40520-025-03073-w

RESEARCH

A head-to-head comparison of the validity and predictive ability for health outcomes of diagnosis versus medication-based comorbidity indices

Moritz Platen[1] · Maresa Buchholz[1] · Anika Rädke[1] · Fabian Kleinke[2] · Annelie Scharf[1] · Michelle Pfaff[1] · Audrey Iskandar[1] · Neeltje van den Berg[2] · Wolfgang Hoffmann[1,2] · Bernhard Michalowsky[1]

Received: 23 September 2024 / Accepted: 12 May 2025
© The Author(s) 2025

Abstract

Background Various comorbidity indices have been validated for individual health outcomes. However, systematic concurrent comparisons of multiple outcome measures in a single study remain relatively underrepresented but needed for practical decision support.

Aims To compare the performance of the Charlson Comorbidity Index (CCI) and Rx-Risk Comorbidity Index (Rx-Risk).

Methods Baseline and six-month follow-up data from $n=221$ patients recruited in $n=70$ practices were used. CCI and Rx-Risk scores were calculated using documented diagnoses and prescribed medications. Outcomes assessed were health-related quality of life (HRQoL, EQ-5D-5 L), functional impairment (B-ADL), cognitive decline (MMSE), and healthcare utilization (physician visits, hospitalizations). Indices performance was evaluated regarding agreement (Cohens Kappa (k)), known-groups validity (ANOVA, t-test), convergent validity (correlation coefficient (r_s)) and predictive ability (R^2, Akaike information criterion (AIC)).

Results Patients were, on average, 80 years old, mostly female (55%), with 12 diagnoses and seven medications. Agreement between both indices was poor for all conditions except diabetes ($k=0.645$) and chronic airway diseases ($k=0.486$). Rx-Risk differed more in known groups, especially for HRQoL and hospitalizations, and showed stronger correlations with the EQ-5D index (r_s, -0.215 vs. -0.134) and risk of hospitalization (r_s, 0.145 vs. 0.128) than CCI. Rx-Risk, again, performed better in predicting the change of EQ-5D index (R^2, 30 vs. 28%) and all EQ-5D dimensions, functional (R^2, 55 vs. 52%) and cognitive decline (R^2, 47 vs. 46%) and physician consultations (AIC, 649.2 vs. 651.0), except for hospitalization (AIC, 149.2 vs. 147.1).

Conclusions Rx-Risk demonstrated slightly superior validity and predictive ability for HRQoL and healthcare utilization, making it a promising option for studies focused on these outcomes. However, limitations regarding functional and cognitive impairment suggest alternative instruments are needed.

Keywords Multimorbidity · Comorbidity indices · Risk adjustment · Patient-reported outcomes · Predictive models

Introduction

Multimorbidity or comorbidity is highly prevalent in the older population [1, 2] and is often associated with poor health-related quality of life (HRQoL) [3], mortality [4], loss of physical function [5], and adverse events such as hospitalizations [6], which also has health economic implications. Several comorbidity indices were developed to quantify the specific burden of chronic conditions, control for confounding and adjust for risks comparing health outcomes or predicting health changes [7–9]. However, there are differences

✉ Moritz Platen
moritz.platen@dzne.de

[1] German Center for Neurodegenerative Diseases (DZNE), site Rostock/Greifswald, Ellernholzstrasse 1-2, D-17489 Greifswald, Germany

[2] Institute for Community Medicine, Section Epidemiology of Health Care and Community Health, University Medicine Greifswald (UMG), Ellernholzstrasse 1-2, D-17489 Greifswald, Germany

Published online: 27 May 2025

in index performance across health outcomes and studies comparing them systematically and concurrently are remain relatively underrepresented.

Generally, a distinction is made between self-reported (e.g. Self-Administered Comorbidity Questionnaire [10]), diagnosis (e.g. Elixhauser comorbidity index [11]) and medication-based (e.g. Chronic Disease Score [12]) comorbidity indices. Self-reported multimorbidity measures perform better regarding HRQoL and function-related outcomes but simultaneously could be affected by patients' subjective perception and recall bias [13–15]. Diagnosis-based metrics are better predictors of mortality and are characterized by easy operationalization due to the availability of data from medical records or reimbursement data but can be affected by coding biases and blind spots regarding the severity of the respective conditions [16–18]. Medication-based measures have been developed to overcome these deficits, providing reliable data due to their more systematic linkage to the underlying diseases and their better performance in predicting health outcomes [18, 19].

Several systematic reviews support choosing the best comorbidity index that fits the designs and methodologic needs [9, 19, 20]. But even though prior research has explored various comorbidity indices there is substantial heterogeneity and significant discrepancies in their performance, underlining the need for standardization in comorbidity measurement to gain more comparable results [4, 19, 21]. Additionally, large-scale studies increasingly aim to capture the multidimensionality of health by examining multiple outcomes– such as clinical, patient-reported, and health-economic measures– simultaneously within a single study design. However, previous validation studies do not take into account the aspect of examining several health-related outcomes within the same patient population. In particular, outcomes like subjective health perspectives (HRQoL), functional impairment regarding activities of daily living, cognitive decline, and healthcare utilization (e.g., physician visits, hospitalizations) provide complementary information on older adults' health. Although established comorbidity indices aim at the holistic quantification of the chronic disease burden, it is difficult to adequately capture the interplay between physical, cognitive, and psychosocial factors. This poses specific methodical challenges and often conflicts with ambitions such as data minimization. Therefore, research is needed to systematically compare comorbidity indices across concurrent outcomes to guide researchers in selecting the most suitable comorbidity index for their studies.

Thus, the present study aims to evaluate the agreement, known-groups validity, convergent validity and predictive ability of the most frequently used diagnosis- (i.e. Charlson comorbidity index (CCI) [22]) and medication-based (i.e. Rx-Risk comorbidity index (Rx-Risk) [23]) index across multiple complementary outcomes reflecting clinical (functional impairment, cognitive decline), patient-reported (HRQoL) and health-economic (healthcare utilization) domains to provide decision support for future study conceptualization.

Materials and methods

Data source and study population

The present analysis was based on data from the InDePendent trial [24]. Patients were recruited in $n=70$ physician practices who are members of one of five participating physician networks in Germany. A total of $n=471$ patients met the eligibility criteria (DemTect <9 or formally diagnosed with dementia, ≥ 70 years old, living at home) and provided written informed consent. The InDePendent trial was approved by several ethics committees (University Medical Center Greifswald (BB 144/20); the State Medical Council Brandenburg (AS 81(bB)/2020); the State Medical Council of Hesse (2020-2081-zvBO)). The recruitment started in January 2021 and ended in December 2022. The detailed design has been described elsewhere [24].

A total of $n=417$ patients started the baseline assessment, and $n=54$ dropped out after baseline, resulting in a sample of $n=363$ patients who started the 6-month follow-up. Patients who completed the follow-up assessment were considered for the present analyses, resulting in a sample of $n=221$ patients.

Data assessment

Nurses conducted a comprehensive, standardized, computer-assisted interview at the participants' homes at baseline and six months after, assessing patients' sociodemographics (age, sex, living situation) and clinical variables. Additionally, ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th revision) diagnoses listed in the general practitioner (GP) files were documented [25], and medication data comprising all regularly taken drugs were validated with medication lists provided by the treating GP or, if available, by the administering nursing service.

Comorbidity scores

Both Charlson and Rx-Risk scores, which represent fundamentally different approaches to measuring multimorbidity for each patient were calculated using documented diagnoses and recorded medications obtained from the medical

records of the treating GP. The currency of diagnoses and regular use of medications at the time of data collection (only baseline) were confirmed by GPs or the drug-administrating nursing services. The CCI was applied, representing the most commonly used diagnose-based comorbidity score considering 17 diseases (Supplementary Tables 1a) [22]. An updated coding algorithm based on the ICD-10 System for the CCI by Quan et al. [26] was used to calculate a weighted sum score, considering the presence of both the respective conditions and their severity. Comorbidity was further categorized as follows according to previous research [27]: no (score < 1), low (score = 1), high (score = 2–3) and very high (score > 3). As a proxy for the medication-based comorbidity index, the Rx-Risk was used comprising 46 comorbidities linked to their indicative Anatomical Therapeutic Chemical (ATC) codes and weighted differently based on their predictive validity for mortality in an elderly population to calculate an overall comorbidity score (Supplementary Tables 1b) [23]. The higher the score, the higher the degree of comorbidity. The Rx-Risk score was then divided into quartiles corresponding to the following categories in line with recent research [28]: no, low, high, and very high degree of comorbidity.

Outcome measures

Health-related quality of life

The EQ-5D-5 L was collected to measure HRQoL [29, 30]. The widely used, generic, preference-based, multidimensional EQ-5D-5 L comprises five dimensions (mobility, self-care, pain/discomfort, usual activities, and anxiety/depression) and five levels (no, slight, moderate, severe, and extreme problems). Additionally, the level sum score (LSS) were calculated and a visual analog scale (EQ-VAS) anchored at 0 (worst health) and 100 (best health) assessed self-reported HRQoL [29–31]. The EQ-5D-5 L responses were converted to health utility values using the German value set, anchored at 0 (death) and 1 (full health) [32].

Functional and cognitive impairment

Functional impairment was measured by the Bayer Activities of Daily Living Scale (B-ADL) [33], ranging from 1 (best functional status) to 10 (worst functional status). Cognitive function was assessed using the Mini-Mental State Examination (MMSE) [34], ranging from 0 to 30, whereas a higher score indicates better cognitive function.

Health resource use

Health resource use was assessed using the Questionnaire for Health-related Resources in Older People (FIMA) [35]. Detailed information about frequencies or quantities of used medical and formal care services was recorded during the mentioned standardized, computer-assisted interviews at the participants' homes at baseline and six months after. Caregivers' proxy ratings were captured to improve data validity and precision. Hospitalization in the past 6 months was assessed dichotomously (at least one vs. none), and GP use in the previous quarter was assessed as a count data (number of consultations).

Statistical analyses

All statistical analyses were conducted with STATA/IC software, version 16 [36].

Agreement

The agreement between both indices for overlapping conditions (congestive heart failure, diabetes, gastric condition, liver disease, chronic airway disease, dementia, renal disease) was assessed by Cohens Kappa (k), whereby $k > 0.40$ indicates acceptable agreement [37].

Known-groups validity

Known-groups validity was evaluated to distinguish between the degree of comorbidity and different stages of HRQoL (EQ-5D-5 L index, dimensions, VAS), functional (B-ADL) and cognitive (MMSE) impairment, and health resource use (GP contacts and hospitalization) using one-way analyses of variance (> 2 groups) and t-tests (2 groups) or Pearson's chi-squared test (sensitivity analyses). Moreover, different comorbidity categories were compared regarding the mentioned outcomes for sensitivity analyses.

Convergent validity

Convergent validity between the CCI and Rx-Risk scores and outcome variables was assessed using the Pearson correlation coefficient (r_p, parametric variables), Spearman's correlation coefficient (r_s, nonparametric variables) and point-biserial correlation coefficient (r_{pb}, binary variables). A correlation coefficient > 0.3 was determined as a moderate correlation [38].

Table 1 Patient characteristics

	Total sample ($n=221$)	
Demographics		
Age, mean (SD)	79.9	(6.5)
Sex (female), n (%)	122	(55.2)
Living alone, n (%)	93	(42.1)
Clinical characteristics		
Cognitive status (MMSE), mean (SD)	18.6	(6.9)
Depression (GDS), mean (SD)	3.3	(2.7)
Functional impairment (B-ADL), mean (SD)	5.6	(2.2)
Number of diagnoses, mean (SD)	11.8	(10.2)
Charlson score, mean (SD)	2.4	(2.6)
Number of medications, mean (SD)	6.8	(3.6)
Rx-Risk score, mean (SD)	3.5	(3.4)

Abbreviations: *B-ADL* Bayer-Activities of Daily Living Scale, range 0–10, lower score indicates better performance; *GDS* Geriatric Depression Scale, sum score 0–15, score≥6 indicates depression; *ICD* International Statistical Classification of Diseases and Related Health Problems; *MMSE* Mini-Mental State Examination, range 0–30, higher score indicates better cognitive function; *SD* standard deviation

Predictive ability

Multiple regression models were performed with the respective comorbidity index as a predictor of interest and each outcome six months after baseline as a dependent variable to test the predictive ability. Specifically, for continuous outcomes (EQ-5D-5 L VAS, index, LSS, BADL, and MMSE), general linear regression models were conducted. For the ordinal outcome (EQ-5D-5 L dimensions), ordinal logistic regression models were used. The count outcome (GP contacts) was analyzed using Poisson regression models, and the binary outcome (hospitalizations) was assessed using logistic regression models. Models were adjusted for age, sex, living situation, and the baseline value of each outcome and standard errors were clustered at the GP practice level. The predictive performance of the indices was determined by the coefficient R^2 for parametric variables or the Akaike information criterion (AIC) for nonparametric variables [39]. A higher R^2 and lower AIC indicate better model performances [39]. The evidence for a better fitting model is determined by AIC differences between the compared models as follows: 0–2 (weak), 2–6 (positive), 6–10 (strong) and >10 (very strong) [40]. Furthermore, the receiver operating characteristic (ROC) curves and c statistic were computed in addition to the AIC for the risk of hospitalization. C statistics>0.8 indicate excellent discrimination [41].

Results

Sample characteristics

Table 1 summarizes the patient characteristics.

Agreement

The analysis revealed poor agreements between overlapped conditions, indicated by $k<0.4$, except for diabetes ($k=0.645$, CI 95% 0.529–0.762) and chronic airway diseases ($k=0.486$, CI 95% 0.300–0.672). The CCI covers more identified cases than the Rx-Risk, particularly for diabetes (81% vs. 78%), chronic airways (69% vs. 46%), dementia (92% vs. 37%), and renal diseases (100% vs. 3%), but was inferior to the Rx-Risk for congestive heart failure (17% vs. 98%), gastric conditions (6% vs. 97%) and liver diseases (29% vs. 71%) (see Table 2).

Known-groups validity

The CCI could not statistically significantly differentiate between the known groups. In contrast, Rx-Risk could distinguish between the tertiles of the EQ-5D index score, the VAS, and the different stages of problems encountered in mobility and anxiety/depression or hospitalizations. Different age groups, further EQ-5D dimensions, cognitive and functional impairment, and GP consultations could be differentiated neither by the CCI nor the Rx-Risk index

Table 2 Comparison of CCI and Rx-Risk in identifying medical conditions (Agreement)

	Patients identified by						
	at least one index *n* (%)	both indices *n* (%)	CCI only *n* (%)	RxRisk Index only *n* (%)	CCI *n* (%)	RxRisk Index *n* (%)	k (95% CI)
Congestive heart failure	118 (100)	18 (15.3)	2 (1.7)	98 (83.1)	20 (17.0)	116 (98.3)	0.131 (0.060–0.201)
Diabetes	**72 (100)**	**42 (62.2)**	**16 (22.2)**	**14 (19.4)**	**58 (80.6)**	**56 (77.8)**	**0.645 (0.529–0.762)**
Gastric condition	73 (100)	2 (2.7)	2 (2.7)	69 (94.5)	4 (5.5)	71 (97.3)	0.020 (-0.037–0.077)
Liver disease	7 (100)	N/A	2 (28.6)	5 (71.4)	2 (28.6)	5 (71.4)	-0.013 (-0.027–0.0004)
Chronic airways disease	**35 (100)**	**13 (37.1)**	**11 (45.8)**	**11 (45.8)**	**24 (68.6)**	**24 (68.6)**	**0.486 (0.300–0.672)**
Dementia	160 (100)	50 (31.3)	97 (66.0)	13 (8.1)	147 (91.9)	63 (37.4)	0.128 (0.035–0.222)
Renal disease	39 (100)	1 (2.6)	38 (97.4)	1 (2.6)	39 (100)	1 (2.6)	0.042 (-0.038–0.1213)

Abbreviations: *CCI* Charlson comorbidity index; *CI* confidence interval; *k* Cohens Kappa; *Rx-Risk* Rx-Risk comorbidity index

Bold values indicate k>0.40; acceptable agreement

(see Table 3). Sensitivity analyses using different levels of comorbidity of the CCI and Rx-Risk index could significantly discriminate the EQ-5D index score or LSS, respectively (see Supplementary Table 2).

Convergent validity

The Rx-Risk demonstrated more often statistically significant ($p<0.05$) correlations with the outcome measures and larger correlation coefficients than the CCI, especially with the EQ-5D index (*rp* -0.215 vs. -0.134) and the LSS (*rp* 0.194 vs. 0.115), several EQ-5D dimensions, such as pain and discomfort (*rs* 0.170 vs. 0.165), anxiety and depression (*rs* 0.144 vs. 0.073), and the risk of hospitalization (*rpb* 0.145 vs. 0.128). Only for the EQ-5D dimension of mobility is the Rx-Risk score inferior to the CCI score (*rs* 0.142 vs. 0.172). However, in line with the definition of a moderate correlation ($r>0.3$), none of the correlation coefficients reached this threshold. There was no correlation between indices and the EQ-5D VAS, EQ-5D dimensions of self-care, usual activities, cognitive and functional impairment, or GP consultations (see Table 4).

Predictive ability

The Rx-Risk is superior to the CCI in predicting the proportion of variation (R^2) for the association between baseline and the six-month follow-up on the EQ-5D index (R^2, 30 vs. 28%), LSS (R^2, 36 vs. 35%), and functional (R^2, 55 vs. 52%) and cognitive impairment (R^2, 47 vs. 46%). The previously defined threshold values for AIC differences ranged from weak (0–2) to positive (2–6). The Rx-Risk score performed better than the CCI in predicting all EQ-5D dimensions (mobility (AIC, 489.4 vs. 490.2); self-care (AIC, 552.5 vs. 557.5); usual activities (AIC, 580.6 vs. 584.4); pain/discomfort (AIC, 467.2 vs. 470.0); anxiety/depression (AIC, 446.9 vs. 447.3)) and GP consultations (AIC, 649.2 vs. 651.0). In contrast; the Rx-Risk was inferior to the CCI regarding the EQ-5D VAS (R^2, 9% vs. 10%) and risk of hospitalization (AIC, 149.2 vs. 147.1). Additionally, the c statistic for risk of hospitalization indicated that both indices exhibited excellent (i.e. c statistics > 0.8) discrimination (ROC area, CCI: 0.811 vs. Rx-Risk: 0.817) (see Table 5; Fig. 1).

Discussion

This study confirmed that diagnosis-based (CCI) and medication-based (Rx-Risk) indices could tremendously differ in capturing chronic conditions, overlapping in only six chronic conditions with mainly poor agreement. Overall, the Rx-Risk was superior to the CCI in terms of known-groups validity and slightly superior regarding convergent validity and predictive ability. Differences favouring the CCI were insufficient to outperform the Rx-Risk noticeably.

Previously published studies have shown that the medication- and diagnosis-based indices differ significantly in detecting diseases covered by both, indicating a general poor agreement except for diabetes, which aligns with the present findings [42–44]. Biases and underestimation may accompany this lack of agreement. For dementia, the CCI detects 92%, and the Rx-Risk only 37% of cases identified as the sum of both instruments, indicating an underestimation of cases using the Rx-Risk in the German population. The underestimation problem becomes more evident, considering that even the CCI identified only 67% of the patients included in the present study, addressing the problem of underdiagnosing dementia in primary care, which should be considered when choosing the right instrument depending on specific populations to avoid potential biases [45, 46].

The Rx-Risk significantly outperforms the CCI in terms of the known-groups validity. Known groups of HRQoL (EQ-5D-5 L index score and VAS terciles), of experienced problems in the dimensions of mobility and anxiety/depression, and of hospitalizations differed considerably by the Rx-Risk index score. Importantly, when looking at the observed changes measured by the EQ-5D-5 L, mean changes particularly of the EQ-5D-5 L index terciles were higher than the minimal clinically important differences of 0.05 in chronic conditions [47]. However, known groups of age, other EQ-5D dimensions, cognitive and functional impairments, and GP visits could not be considerably delimited. Although the CCI is commonly applied to various outcomes in health services and clinical research, its inferior discriminative ability regarding these outcomes could be explained by the fact that it was initially conceptualized to predict mortality. In contrast, the Rx-Risk was updated and adjusted also for health economic outcomes, such as healthcare costs [19]. Applying both indices to the cognitive and functional impairment could be misleading. Although the indices cover multiple conditions that can potentially impact cognitive and functional decline, the instruments used could not reflect the different stages of dementia, which should be considered for trials conceptualization.

Regarding the convergent validity, several studies underline comparatively poor correlations between the respective instruments, especially HRQoL, and a slight inferiority of the CCI compared to medication-based scores [18, 48]. Consequently, these studies advise against its use for research examining HRQoL. This aligns with the present analysis, indicating only low correlations and a slight superiority of the Rx-Risk instrument over the CCI for HRQoL. Additionally, Beloosesky et al. [49] found at least correlations for

Table 3 Comparing known-groups validity of CCI and Rx-Risk

	Comorbidity Index, $n=221$			p value
	CCI (score) mean (SD)	p value	Rx-Risk (score) mean (SD)	
Age				
≤81 years old, $n=118$	2.45 (2.84)	0.927	3.74 (3.31)	0.324
>81 years old, $n=103$	2.42 (2.23)		3.28 (3.54)	
Health-related Quality of Life				
EQ-5D-5 L VAS				
1st tercile (≤51), $n=75$	2.67 (2.36)	0.450	4.44 (3.80)	**0.017**
2nd tercile (52–70), $n=90$	2.18 (2.56)		3.06 (3.15)	
3rd tercile (≥71), $n=56$	2.54 (2.83)		3.05 (3.09)	
EQ-5D-5 L Index score				
1st tercile (≤0.771), $n=74$	2.81 (2.45)	0.271	4.35 (3.58)	**0.038**
2nd tercile (0.772–0.887), $n=75$	2.15 (2.57)		3.09 (3.33)	
3rd tercile (≥0.888), $n=72$	2.35 (2.67)		3.13 (3.23)	
EQ-5D-5 L Dimensions				
Mobility				
No problem, $n=64$	2.14 (2.61)	0.078	3.27 (2.88)	**≤0.001**
Slight to moderate problems, $n=131$	2.37 (2.46)		3.11 (3.42)	
Severe to extreme problems, $n=26$	3.46 (2.79)		6.27 (3.50)	
Self-care				
No problem, $n=90$	2.58 (2.77)	0.113	3.40 (3.37)	0.154
Slight to moderate problems, $n=109$	2.14 (2.41)		3.36 (3.33)	
Severe to extreme problems, $n=22$	3.32 (2.30)		4.86 (3.89)	
Usual activities				
No problem, $n=56$	2.55 (3.00)	0.497	3.52 (3.64)	0.333
Slight to moderate problems, $n=135$	2.29 (2.42)		3.34 (3.38)	
Severe to extreme problems, $n=30$	2.87 (2.32)		4.37 (3.16)	
Pain/ discomfort				
No problem, $n=83$	2.11 (2.19)	0.065	2.93 (3.36)	0.086
Slight to moderate problems, $n=126$	2.51 (2.71)		3.80 (3.39)	
Severe to extreme problems, $n=12$	3.92 (3.03)		4.75 (3.79)	
Anxiety/depression				
No problem, $n=115$	2.41 (2.92)	0.987	3.22 (3.52)	**0.032**
Slight to moderate problems, $n=94$	2.46 (2.19)		3.60 (3.12)	
Severe to extreme problems, $n=12$	2.50 (1.68)		5.92 (4.08)	
Functional impairment (B-ADL)				
No problems, $n=41$	2.85 (3.16)	0.424	3.05 (3.29)	0.507
Moderate problems, $n=81$	2.47 (2.57)		3.46 (2.94)	
Severe problems, $n=99$	2.23 (2.28)		3.78 (3.83)	
Cognitive status (MMSE)				
No hint for dementia, $n=20$	3.20 (2.91)	0.145	3.00 (2.49)	0.760
Mild dementia, $n=98$	2.10 (2.14)		3.53 (3.52)	
Moderate/Severe dementia, $n=103$	2.60 (2.83)		3.62 (3.50)	
Health resource utilization				
GP consultations				
1 visit, $n=96$	2.30 (2.71)	0.299	3.67 (3.44)	0.067
2 visits, $n=58$	2.19 (2.11)		2.67 (3.10)	
≥3 visits, $n=67$	2.84 (2.68)		4.06 (3.56)	
Hospitalization				
Yes, $n=50$	3.04 (3.27)	0.057	4.44 (3.69)	**0.031**
No, $n=171$	2.26 (2.30)		3.26 (3.30)	

Abbreviations: *B-ADL* Bayer-Activities of Daily Living Scale; *CCI* Charlson comorbidity index; *GP* general practitioner; *MMSE* Mini-Mental State Examination; *r* correlation coefficient; *Rx-Risk* Rx-Risk comorbidity index; *VAS* visual analog scale

Table 4 Comparing convergent validity of CCI and Rx-Risk

	CCI (score), n=221		RxRisk (score), n=221	
	r	p value	*r*	p value
Health-related Quality of Life				
EQ-5D-5L VAS	-0.026[1]	0.698	-0.130[1]	0.053
EQ-5D-5L index score	**-0.134**[1]	**0.047**	**-0.215**[1]	**0.001**
EQ-5D-5L Level sum score	0.115[1]	0.087	**0.194**[1]	**0.004**
EQ-5D-5L Dimensions				
Mobility	**0.172**[2]	**0.010**	**0.142**[2]	**0.035**
Self-care	0.040[2]	0.554	0.093[2]	0.171
Usual activities	0.094[2]	0.165	0.129[2]	0.055
Pain/ discomfort	**0.165**[2]	**0.014**	**0.170**[2]	**0.012**
Anxiety/depression	0.073[2]	0.283	**0.144**[2]	**0.032**
Functional impairment (B-ADL)	-0.084[1]	0.215	0.097[1]	0.151
Cognitive status (MMSE)	0.036[1]	0.596	0.033[1]	0.631
Health resource utilization				
GP consultations	0.028[1]	0.675	-0.035[1]	0.605
Hospitalization	0.128[3]	0.058	**0.145**[3]	**0.032**

Abbreviations: *B-ADL* Bayer-Activities of Daily Living Scale; *CCI* Charlson comorbidity index; *GP* general practitioner; *MMSE* Mini-Mental State Examination; *r* correlation coefficient; *Rx-Risk* Rx-Risk comorbidity index; *VAS* visual analog scale

[1] Pearson correlation coefficient [2] Spearman's correlation coefficient [3] Point-biserial correlation coefficient

Bold values indicate P< 0.05

shadings indicate superiority

Table 5 Comparing the predictive ability of CCI and Rx-Risk

Health outcomes	Charlson Comorbidity Index (score), n=221			RxRisk Index (score), n=221		
Health-related Quality of Life	**R^2**	**p value (model)**	**B (SE)**	**R^2**	**p value (model)**	**B (SE)**
EQ-5D-5L VAS	0.10	0.009	0.40 (0.43)	0.09	0.010	-0.25 (0.34)
EQ-5D-5L index score	0.28	≤0.001	-0.001 (0.004)	0.30	≤0.001	-0.008 (0.003)*
EQ-5D-5L Level sum score	0.34	≤0.001	0.075 (0.08)	0.36	≤0.001	0.14 (0.06)*
EQ-5D-5L Dimensions	**AIC**	**p value (model)**	**OR (SE)**	**AIC**	**p value (model)**	**OR (SE)**
Mobility	490.2	≤0.001	0.98 (0.06)	489.4	≤0.001	1.04 (0.05)
Self-care	557.5	≤0.001	1.11 (0.06)*	552.5	≤0.001	1.13 (0.04)**
Usual activities	584.4	≤0.001	1.08 (0.06)	580.6	≤0.001	1.11 (0.05)*
Pain/ discomfort	470.0	≤0.001	1.01 (0.06)	467.2	≤0.001	1.08 (0.05)
Anxiety/depression	447.3	≤0.001	0.97 (0.06)	446.9	≤0.001	1.03 (0.04)
	R^2	**p value (model)**	**B (SE)**	**R^2**	**p value (model)**	**B (SE)**
Functional impairment (B-ADL)	0.52	≤0.001	0.04 (0.04)	0.55	≤0.001	0.11 (0.03)***
Cognitive status (MMSE)	0.46	≤0.001	-0.18 (0.16)	0.47	≤0.001	-0.22 (0.12)
Health resource utilization						
	AIC	**p value (model)**	**B (SE)**	**AIC**	**p value (model)**	**B (SE)**
GP consultations	651.0	0.057	0.01 (0.01)	649.2	≤0.001	0.02 (0.01)*
	AIC	**p value (model)**	**OR (SE)**	**AIC**	**p value (model)**	**OR (SE)**
Hospitalization yes	147.1	≤0.001	1.17 (0.10)	149.2	≤0.001	1.09 (0.08)

Abbreviations: *AIC* Akaike information criterion; B observed coefficient; *B-ADL* Bayer-Activities of Daily Living Scale; *GP* general practitioner; *MMSE* Mini-Mental State Examination; *OR* OR odds ratio; *SE* standard error; *VAS* visual analog scale

Separate models were fitted for each health outcome and index, respectively. used were adjusted for age, sex, and living situation, and the respective baseline value of the dependent variables.

p value (coefficient) < 0.05*, p value < 0.01**, p value < 0.001***

shadings indicate superiority

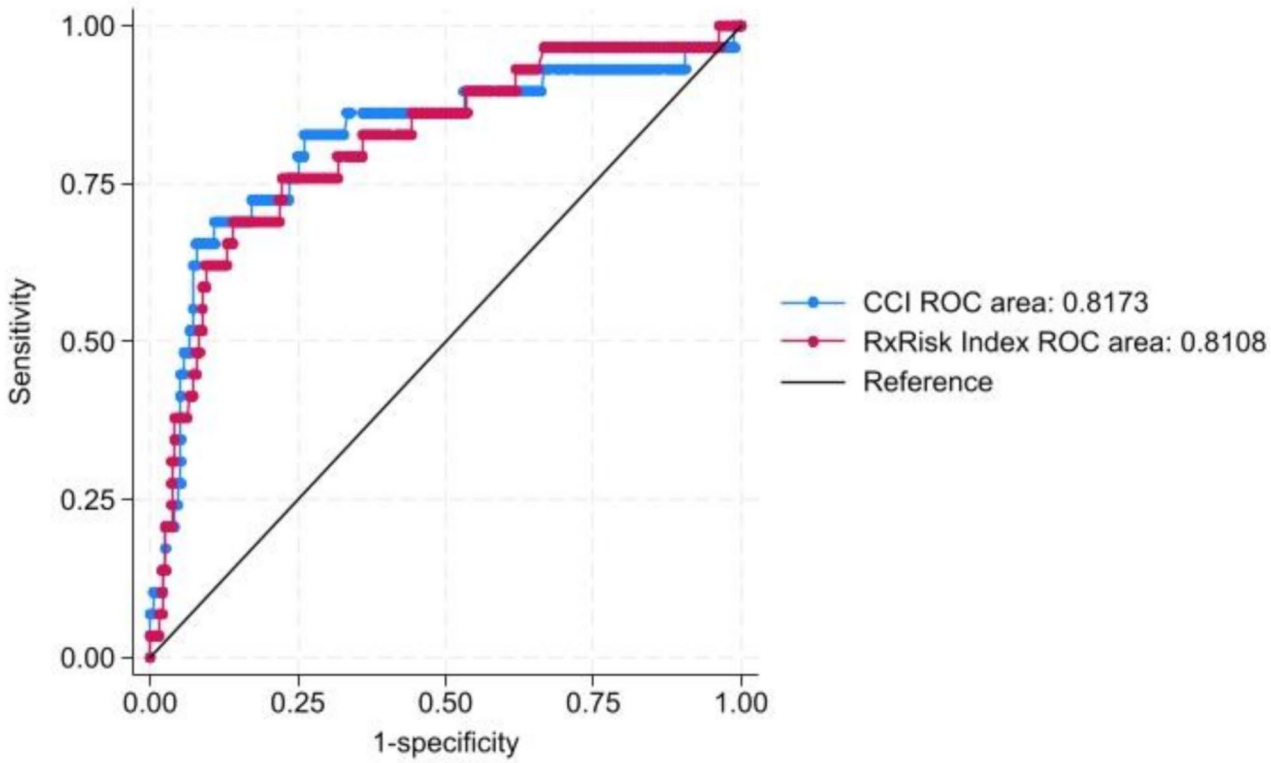


Fig. 1 Comparing CCI and Rx-Risk using ROC Curves for predictability of hospitalization
Abbreviations: *CCI* Charlson comorbidity index; *ROC* Receiver operating characteristics; *Rx-Risk* Rx-Risk comorbidity index

the CCI with functional and cognitive impairment, whereas neither index exhibits significant results in the present analysis. However, participants were less cognitively and functionally impaired than in the present sample, which may suggest saturation effects, indicating insufficient variations at advanced levels of impairment and making additional effects unlikely to be observed [49]. Since none of the correlation coefficients reached the threshold of > 0.3 for moderate correlation, indicating overall poor correlations the practical implications of these correlations should be interpreted with caution, as the small differences in correlation coefficients may not translate into meaningful differences in patient care. The findings underline that the choice of the proper instrument depends not only on the target outcomes but also on the characteristics of the targeted patient populations. Therefore, using both indices for aged, functionally, and cognitively impaired patients is not recommended.

The multivariate analyses, predicting future values of the respective outcomes, confirmed the slight superiority of the medication-based Rx-Risk index compared to the diagnosis-based CCI. It is important to emphasize that despite the statistical superiority, the observed differences are relatively small; hence, the choice between indices should also consider contextual and practical aspects beyond numerical performance alone. Although both indices were initially not targeted on HRQoL, both were significant predictors for HRQoL change, with a slightly better predictive ability of the Rx-Risk index. The inferiority of the CCI was expected and underlined by several studies evaluating the performance of various indices in predicting HRQoL [48, 50, 51]. Additionally, Sasseville et al. [52] found a higher specificity of the medication-based compared to the diagnosis-based index, demonstrating poorer HRQoL according to the EQ-5D for increased multimorbidity. Despite the present findings indicating the suitability of the Rx-Risk for predicting HRQoL changes, further evidence from validation in appropriate populations is required for a final assessment and recommendation for use.

While the present study shows promising results for both instruments in predicting functional and cognitive impairment, with a slight superiority of the Rx-Risk instrument, recently published studies have shown that medication- and diagnosis-based indices, including the Rx-Risk and CCI, are either equivalent or show only poor discrimination, predicting especially functional impairment [52, 53]. Thus, the present results should also be used cautiously. Modified and validated instruments adapted to patients with cognitive impairment have been suggested to better predict cognitive status (e.g. the RxDx-Dementia Risk Index [54]).

In contrast to HRQoL, functional and cognitive impairment, in our study, the CCI performed better than the Rx-Risk regarding hospitalizations, which is inconsistent with previously published evidence. Wallace et al. [53] compared five diagnosis- and medication-based indices and found that the discrimination for all measures was poor, and all instruments were equivalent in predictive ability. However, the present results show excellent discrimination between patients at higher risk for hospitalizations and those at lower risk by Rx-Risk and CCI. Dominick et al. [55] exhibited the superiority of the Rx-Risk towards the CCI for physician visits and hospitalization in osteoarthritis patients. While the present results for physician visits were confirmed, those for hospitalizations were the opposite. Several studies suggest uncertainties in various instruments to predict hospitalizations, as emphasized by Radomski et al. [56], elaborating that the different comorbidity indices could drastically affect the direction of effect in outcomes. Additionally, Borson et al. [21] pointed out that different instruments adjusted for comorbidity are not equivalent and that dementia magnifies these discrepancies, which could also apply to the present sample. A standardized, valid, and reliable instrument adapted to dementia is needed to overcome these uncertainties.

Limitations

Data were obtained from a small sample of German community-dwelling primary care patients, potentially limiting the generalizability of the presented results for in-patient or institutionalized patients. Participants included could therefore tend to have a lower degree of comorbidity. Moreover, the small sample size increases the risk of a Type II error (i.e., insufficient statistical power to detect true effects), which may limit the interpretation of the results. However, there are no standard criteria for calculating the sample size for this type of study. Concerning the psychometric validation, Anthoine et al. [57] revealed that the median sample size of 114 randomly selected validation studies was 207. As a rule of thumb, Rouquette et al. [58] demonstrated that a minimum of 300 subjects is generally acceptable for a validation study. The present study is based on a sample size of 221 patients. Therefore, a post hoc power analysis (Supplementary Table 3) was conducted for correlation and regression analyses indicating insufficient statistical power to detect true effects regarding the correlation analysis or convergent validity for all outcomes except for the EQ-5D-5 L index score and LSS. In contrast, the post hoc power analysis for the models that assessed the predictive ability achieved sufficient power for each outcome except for the number of GP consultations. Nevertheless, future studies with larger, more diverse populations are necessary to confirm and extend the present findings and to ensure sufficient power in discriminating relevant outcomes. Patients without formal dementia diagnosis at the time of study enrollment were screened for dementia, which could lead to false-positive screening results and consequently to a formal diagnosis by the treating GP that may not be valid in all cases. Both examined comorbidity indices, the Rx-Risk and CCI, were initially targeted to predict outcomes in administrative data, which could reflect coding and accounting behaviour; thus, results derived from their use in primary collected data may differ. Furthermore, both indices assume that patients' diseases are recognized or treated and that they therefore have access to healthcare, which can lead to underestimation and the mentioned bias in coding. Additionally, neither index has been validated specifically for the German population.

Conclusion

Comparing the diagnosis-based CCI and medication-based Rx-Risk indices revealed tremendous differences. The two indices captured different chronic conditions with only some overlapping conditions and could not be used as mutual substitutes for each other. However, the Rx-Risk was superior in differentiating between known groups and predictive ability. The Rx-Risk index mainly performed better in predicting HRQoL changes over time.

However, caution is warranted due to the relatively small sample size, which may indicate insufficient statistical power to detect true effects. Particularly discrepancies in predicting health resource utilization and the limitations in assessing functional and cognitive impairment underscore the need for further research based on larger sample sizes. Alternative instruments may offer better predictions for the patient-reported outcomes, such as functional and cognitive status, and should, therefore, be used as covariates in prediction models instead. Thus, standardized, reliable instruments, particularly adapted to the underlying condition of interest, are needed to overcome uncertainties in predicting outcomes and changes in outcomes. A decision for or against a specific multimorbidity index should always consider the nature and the number of outcomes and the specific population planned for future studies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40520-025-03073-w.

Acknowledgements We would like to thank all participating patients, their caregivers, and the participating general practitioners and specialists for their most valued collaboration.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were

performed by MP, MB, FK and BM. The first draft of the manuscript was written by MP and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.
The Innovation Fund at the Federal Joint Committee (G-BA Grand No: 01NVF18034) funds the study. The funders have not influenced the conceptualization and conduct of the study and will not have any role in the data analysis and publication of the results.

Data availability No datasets were generated or analysed during the current study.

Declarations

Ethics approval The study protocol was approved by several ethics committees (University Medical Center Greifswald (BB 144/20); the State Medical Council Brandenburg (AS 81(bB)/2020); the State Medical Council of Hesse (2020-2081-zvBO)).

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Wallace E et al (2015) Managing patients with Multimorbidity in primary care. BMJ 350:h176
2. Nicholson K et al (2019) Multimorbidity and comorbidity revisited: refining the concepts for international health research. J Clin Epidemiol 105:142–146
3. Nelis SM et al (2019) The impact of co-morbidity on the quality of life of people with dementia: findings from the IDEAL study. Age Ageing 48(3):361–367
4. Nunes BP et al (2016) Multimorbidity and mortality in older adults: A systematic review and meta-analysis. Arch Gerontol Geriatr 67:130–138
5. Ryan A et al (2015) Multimorbidity and functional decline in community-dwelling adults: a systematic review. Health Qual Life Outcomes 13:168
6. Rodrigues LP et al (2022) Association between Multimorbidity and hospitalization in older adults: systematic review and meta-analysis. Age Ageing, 51(7)
7. Fishman PA et al (2003) Risk adjustment using automated ambulatory pharmacy data: the RxRisk model. Med Care 41(1):84–99
8. Schneeweiss S, Maclure M (2000) Use of comorbidity scores for control of confounding in studies using administrative databases. Int J Epidemiol 29(5):891–898
9. Stirland LE et al (2020) Measuring Multimorbidity beyond counting diseases: systematic review of community and population studies and guide to index choice. BMJ 368:m160
10. Sangha O et al (2003) The Self-Administered comorbidity questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum 49(2):156–163
11. Elixhauser A et al (1998) Comorbidity measures for use with administrative data. Med Care 36(1):8–27
12. Von Korff M, Wagner EH, Saunders K (1992) A chronic disease score from automated pharmacy data. J Clin Epidemiol 45(2):197–203
13. Bayliss EA, Ellis JL, Steiner JF (2009) Seniors' self-reported Multimorbidity captured biopsychosocial factors not incorporated into two other data-based morbidity measures. J Clin Epidemiol 62(5):550–7e1
14. Schmier JK, Halpern MT (2004) Patient recall and recall bias of health state and health status. Expert Rev Pharmacoecon Outcomes Res 4(2):159–163
15. Silliman RA, Lash TL (1999) Comparison of interview-based and medical-record based indices of comorbidity among breast cancer patients. Med Care 37(4):339–349
16. Austin SR et al (2015) Why summary comorbidity measures such as the Charlson comorbidity index and elixhauser score work. Med Care 53(9):e65–72
17. Iezzoni LI et al (1992) Comorbidities, complications, and coding bias. Does the number of diagnosis codes matter in predicting in-hospital mortality? JAMA 267(16):2197–2203
18. Yurkovich M et al (2015) A systematic review identifies valid comorbidity indices derived from administrative health data. J Clin Epidemiol 68(1):3–14
19. Mehta HB et al (2021) More than two-dozen prescription drug-based risk scores are available for risk adjustment: A systematic review. J Clin Epidemiol 137:113–125
20. Oemrawsingh A et al (2020) Patient-Reported morbidity instruments: A systematic review. Value Health 23(6):791–811
21. Borson S et al (2010) Comorbidity in aging and dementia: scales differ, and the difference matters. Am J Geriatr Psychiatry 18(11):999–1006
22. Charlson ME et al (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 40(5):373–383
23. Pratt NL et al (2018) The validity of the Rx-Risk comorbidity index using medicines mapped to the anatomical therapeutic chemical (ATC) classification system. BMJ Open 8(4):e021122
24. Kleinke F et al (2022) Advanced nursing practice and interprofessional dementia care (InDePendent): study protocol for a multicenter, cluster-randomized, controlled, interventional trial. Trials 23(1):290
25. World Health Organization (WHO), The ICD-10 Classification of Mental and Behavioural Disorders - Diagnostic criteria for research (1993) Geneva
26. Quan H et al (2005) Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 43(11):1130–1139
27. Kaczynski A et al (2019) Comorbidity in dementia diseases and associated health care resources utilization and cost. J Alzheimers Dis 68(2):635–646
28. Kristensen PK et al (2023) Current trends in comorbidity prevalence and associated mortality in a Population-Based cohort of hip fracture patients in Denmark. Clin Epidemiol 15:839–853
29. EuroQol–a new facility for the measurement of health-related quality of life. Health Policy, (1990) 16(3): p. 199–208
30. Devlin NJ, Krabbe PF (2013) The development of new research methods for the valuation of EQ-5D-5L. Eur J Health Econ 14(Suppl 1):S1–3

31. Rabin R, de Charro F (2001) EQ-5D: a measure of health status from the EuroQol group. Ann Med 33(5):337–343
32. Ludwig K, Graf JM, Schulenburg vonder, Greiner W (2018) German value set for the EQ-5D-5L. 36(6):663–674Pharmacoeconomics
33. Hindmarch I et al (1998) The bayer activities of daily living scale (B-ADL). Dement Geriatr Cogn Disord 9(Suppl 2):20–26
34. Kessler J, Markowitsch H, Denzler P (1990) *Mini-Mental-Status-Test (MMST) [German Version]*. 1 ed. Göttingen: Beltz Test GmbH
35. Seidl H et al (2015) FIMA– Fragebogen Zur Erhebung von gesundheitsleistungen Im alter: entwicklung und pilotstudie. Gesundheitswesen 77(01):46–52
36. StataCorp. (2023) *Stata Statistical Software: Release 18. College Station, TX: StataCorp LLC.*
37. Douglas G, Altman (1990) *Practical Statistics for Medical Research*, ed. C.a. Hall/CRC. Vol. 1st Edition. New York
38. Hinkle DE, Wiersma W, Jurs SG (2003) Applied statistics for the behavioral sciences. Houghton Mifflin;, (distributor): Boston, Mass. [London]
39. Bozdogan H (1987) Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. Psychometrika 52(3):345–370
40. Raftery AE (1995) Bayesian model selection in social research. Sociol Methodol 25:111–163
41. David W, Hosmer SL (2000) *Assessing the Fit of the Model*, in *Applied Logistic Regression*. pp. 143–202
42. Inacio MC et al (2015) Comparing co-morbidities in total joint arthroplasty patients using the RxRisk-V, Elixhauser, and Charlson measures: a cross-sectional evaluation. BMC Musculoskelet Disord 16:385
43. Lu CY et al (2011) Charlson and Rx-Risk comorbidity indices were predictive of mortality in the Australian health care setting. J Clin Epidemiol 64(2):223–228
44. Mannion C et al (2020) Agreement between self-reported morbidity and pharmacy claims data for prescribed medications in an older community based population. BMC Geriatr 20(1):283
45. Eichler T et al (2014) Rates of formal diagnosis in people screened positive for dementia in primary care: results of the DelpHi-Trial. J Alzheimers Dis 42(2):451–458
46. Gjøra L et al (2024) Prevalence and determinants of diagnosed dementia: A registry linkage study linking diagnosis of dementia in the Population-Based HUNT study to registry diagnosis of dementia in primary care and hospitals in Norway. J Alzheimers Dis 99(1):363–375
47. Tsiplova K et al (2016) EQ-5D-derived health utilities and minimally important differences for chronic health conditions: 2011 Commonwealth fund survey of sicker adults in Canada. Qual Life Res 25(12):3009–3016
48. Fortin M et al (2005) Comparative assessment of three different indices of Multimorbidity for studies on health-related quality of life. Health Qual Life Outcomes 3:74
49. Beloosesky Y, Weiss A, Mansur N (2011) Validity of the Medication-based disease burden index compared with the Charlson comorbidity index and the cumulative illness rating scale for geriatrics: a cohort study. Drugs Aging 28(12):1007–1014
50. Mehta HB et al (2016) Comparative performance of Diagnosis-based and Prescription-based comorbidity scores to predict Health-related quality of life. Med Care 54(5):519–527
51. Mukherjee B et al (2011) A new comorbidity index: the health-related quality of life comorbidity index. J Clin Epidemiol 64(3):309–319
52. Sasseville M et al (2019) Predicting poorer health outcomes in older community-dwelling patients with Multimorbidity: prospective cohort study assessing the accuracy of different Multimorbidity definitions. BMJ Open 9(1):e023919
53. Wallace E et al (2016) Comparison of count-based Multimorbidity measures in predicting emergency admission and functional decline in older community-dwelling adults: a prospective cohort study. BMJ Open 6(9):e013089
54. Mehta HB et al (2016) Development and validation of the RxDx-Dementia risk index to predict dementia in patients with type 2 diabetes and hypertension. J Alzheimers Dis 49(2):423–432
55. Dominick KL et al (2005) Comparison of three comorbidity measures for predicting health service use in patients with osteoarthritis. Arthritis Rheum 53(5):666–672
56. Radomski TR et al (2017) The impact of Medication-Based risk adjustment on the association between veteran health outcomes and dual health system use. J Gen Intern Med 32(9):967–973
57. Anthoine E et al (2014) Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. Health Qual Life Outcomes 12:176
58. Rouquette A, Falissard B (2011) Sample size requirements for the internal validation of psychiatric scales. Int J Methods Psychiatr Res 20(4):235–249

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.