

1 **Measuring patient centeredness with German language Patient-Reported Experience**
2 **Measures (PREM) – a systematic review and qualitative analysis according to COSMIN**

3

4 André L. Mihaljevic^{1,2*}, Colette Dörr-Harim², Eva Kalkum³, Guido Strunk⁴

5

6 ¹ Department of General and Visceral Medicine, University Hospital Ulm, Albert-Einstein-

7 Allee 23, 89081 Ulm, Germany

8 ² Clinical Trial Centre, Department of Surgery (*ulmCARES*), University Hospital Ulm,

9 Albert-Einstein-Allee 23, 89081 Ulm, Deutschland

10 ³ Study Centre of the German Society of Surgery, Im Neuenheimer Feld 130.3, 69120

11 Heidelberg, Germany

12 ⁴ Complexity-Research, Schönbrunner Str. 32/3/20, A-1050 Wien, Austria

13

14 * corresponding author

15

16 **Corresponding author:**

17 André L. Mihaljevic

18 E-Mail: andre.mihaljevic@uni-ulm.de

19 ORCID iD: 0000-0003-1902-657X

20

21

22 **E-Mail addresses:**

23 Colette Dörr-Harim: colette.doerr-harim@uniklinik-ulm.de

24 Eva Kalkum: eva.kalkum@med.uni-heidelberg.de

25 Guido Strunk: guido.strunk@complexity-research.com

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

26

27 **Category:** Systematic review and qualitative analysis

28 **Word count:** 11.135 (without references), 12.285 (with references)

29 **Key words:** patient-reported experience measure, patient-reported outcome measure, patient-
30 centredness, patient outcome assessment

31

32

33 **Abstract**

34 **Background:** Patient centeredness is an integral part of the quality of care. Patient-reported
35 experience measures (PREMs) are assumed to be an appropriate tool to assess patient-
36 centredness. An evaluation of German-speaking PREMs is lacking.

37 **Objective:** To perform a systematic review and qualitative analysis of psychometric
38 measurement qualities of German-speaking PREMs.

39 **Methods:** A systematic literature search was performed in Medline, PsycInfo, CINHAL,
40 Embase, Cochrane database (last search 9th November 2021) for studies describing generic,
41 surgery- or cancer-specific PREMs. Subsequently, questionnaires that were not in German or
42 for which no German translation exists were excluded. Furthermore, questionnaires, that are
43 for healthcare professionals rather than patients as well as disease-specific questionnaires were
44 excluded. Baseline data for all studies was extracted by two independent reviewers.
45 Psychometric measurement qualities of the questionnaires were assessed according to the
46 COSMIN guidelines.

47 **Results:** In total, 3457 abstracts were screened, of which 3345 were excluded. The remaining
48 112 articles contained 51 PREMs, of which 12 were either developed in or translated into
49 German: 8 generic (NORPEQ, PPE-15, PEACS, HCAHPS, QPPS, DUQUE, PEQ-G,
50 Schoenfelder et al.), 4 cancer-specific (EORTC IN-PATSAT32, PSCC-G, Danish National
51 Cancer Questionnaire, SCCC) and no surgery-specific PREM. None of the PREMs covered all
52 16 domains patient-centeredness. Overall rating of structural validity was adequate only for
53 PEACS and HCAHPS. High ratings for internal consistency were given for NORPEQ,
54 Schoenfelder et al., PSCC-G and the SCCC. Cross-cultural validity for translated
55 questionnaires was adequate only for the PSCC-G, while reliability was adequately assessed
56 only for the EORTC IN-PATSAT32. Due to a lack of measurement gold standard and minimal

57 important change, criterion validity and measurement invariance could not be assessed Most
58 PREMs showed the expected outcomes for hypotheses testing for construct validity.

59 **Conclusion:** None of the PREMs has been fully evaluated in German. Based on current
60 evidence, the EORTC IN-PATSAT32 and PSCC-G can be recommended for cancer patients,
61 while HCAHPS, NORPEQ, PPE-15 and PEACS might be used as generic PREMs. Compared
62 to other languages PREM development in German is rudimentary and no comprehensive
63 PREM system is currently available.

64 **Registration.** PROSPERO CRD42021276827

65 **Funding:** None

66

67 **Introduction**

68 Improving the patient centredness (PC) of healthcare has been a main objective of healthcare
69 politics over the last decades, including German-speaking countries [1,2]. PC has been defined
70 as one of six domains of the quality of care by the Institute of Medicine (IOM), next to safety,
71 effectiveness, timeliness, efficiency and equitability [3] (S1 Fig). However, patients frequently
72 experience a lack of PC in many fields of healthcare [4].

73 Furthermore, the dimensions of PC have not been clearly defined and several models have been
74 proposed in the past (S2 Table). A systematic review has identified 15 dimensions of PC [5]:
75 patient as a unique person, biopsychosocial perspective, essential characteristics of the
76 clinician, patient involvement in care, involvement of family and friends, physical support,
77 emotional support, clinician-patient communication, patient empowerment, patient
78 Information, access to care, integration of medical and non-medical care, coordination and
79 continuity of care, teamwork and teambuilding, clinician-patient relationship. The influential
80 Picker model contains an additional dimension termed “effective treatment by trustworthy and
81 qualified personnel” [6] (S2 Table).

82 Several methods have been proposed to measure PC in clinical practice [7]. Assessment via
83 questionnaires termed *Patient-Reported Experience Measures* (PREM) is most frequent as they
84 permit a standardized appraisal of PC. PREMs aim to measure PC via the experience of patients
85 in a certain healthcare context. Depending on this healthcare context four different categories
86 of PREMs can be distinguished, although they partially overlap:

- 87 a) *generic PREMs* measure general aspects of PC and can be applied across multiple
88 healthcare settings and disciplines.
- 89 b) *discipline-specific PREMs* assess the PC within a certain discipline, e.g. surgery or
90 internal medicine.

91 c) *healthcare pathway-specific PREMs* measure the PC across a specific healthcare
92 pathway e.g., cancer care and thus across disciplines and healthcare settings.

93 d) *disease-specific PREMs* aim to measure the PC of healthcare for patients with a specific
94 disease, e.g., breast cancer.

95

96 According to Bull et al. [8] PREMs (a) enable patients to reflect on their care experience; (b)
97 provide patient-level information to drive service quality improvement; (c) serve as quality
98 indicators for public reporting and benchmarking. PREMs need to be distinguished from
99 *Patient-Reported Outcome Measures* (PROM). PROMs measure medical outcomes, rather
100 than dimensions of PC and therefore, address different domains of healthcare in the IOM
101 quality of care model (S2). Like PROMs, however, PREMs need to have adequate
102 psychometric measurement properties (e.g., validity, reliability) to accurately assess PC.
103 However, non-standardized PREMs or PROMs are frequent [9] and constitute a waste of
104 resources [10]. The COSMIN (COnsensus-based Standards for the selection of health
105 Measurement INstruments) group has defined internationally accepted guidelines for assessing
106 psychometric properties of patient-reported measures [11].

107 Contrary to other languages no comprehensive review is available for German PREMs.
108 Furthermore, the psychometric properties of German PREMs have not been evaluated yet.
109 Therefore, the aim of the study is to perform a systematic review and qualitative analysis of
110 generic, surgery-specific and cancer-care specific German PREMs using the internationally
111 accepted COSMIN guidelines [11].

112

113 **Material and Methods**

114 This systematic review is reported according to current PRISMA guidelines [12]. A PRISMA
115 checklist is attached (S3 Table). The review has been registered (PROSPERO
116 CRD42021276827). No funding has been received for this work. A review protocol had been
117 written prior to data extraction.

118 **Eligibility criteria**

119 Studies that described or analysed the development, validation, or measurement of generic,
120 surgery-specific or cancer healthcare pathway-specific PREMs were included in this
121 systematic review. Included studies needed to be published in a peer-reviewed journal. During
122 the initial search no language limitations were set. The following exclusion criteria were set:

- 123 1. Studies that described general satisfaction questionnaires that did not cover aspects of
124 patient experience.
- 125 2. Studies measuring patient expectation rather than patient experience
- 126 3. Studies that measured single aspects of patient-centeredness and were thus no
127 multidimensional PREMs
- 128 4. Studies describing questionnaires for physicians or proxies and not for patients
129 themselves (i.e., were not patient-reported)
- 130 5. Studies containing PREMs that were not generic, surgery-specific or cancer healthcare
131 pathway-specific. Therefore, PREMs for a specific disease were excluded.

132 **Information sources**

133 The following information sources were searched:

- 134 a. EBM Reviews - Cochrane Database of Systematic Reviews 2005 to November 9, 2021,
135 EBM Reviews - Health Technology Assessment 4th Quarter 2016
- 136 b. Embase 1980 to 2021 Week 40

137 c. Ovid MEDLINE(R) ALL 1946 to November 9, 2021

138 d. APA PsycInfo 1806 to November Week 2 2021

139 e. CINHALL November 9, 2021

140 The last search was performed on 9th November 2021. An existing literature review on PREMs
141 by Bull et al. was used as template [8]. This publication searched PREMs until 31st March
142 2018. Therefore, our search was limited to 2018 till November 2021. The results by Bull et al.
143 were included in this systematic review.

144 **Search**

145 The search algorithm is described in detail in supplement 4 (S4 Text). It is an adaptation of the
146 search algorithm described by Bull et al. [8]. Additional studies were identified by reference
147 searching and full text reading.

148 **Study selection**

149 The references of all generic, surgery-specific or cancer healthcare pathway-specific PREMs
150 were imported into the citation program Zotero (www.zotero.org; Version 5.0.96.2).
151 Duplicates were identified and merged either with the find duplicates function in Zotero or by
152 hand. Titles and abstracts of all articles were read by two reviewers (AMi, CDH) and those
153 studies not fulfilling eligibility criteria were removed. In a next step, the fulltext articles of all
154 remaining studies were read, to decide which articles fulfil eligibility criteria. Fulltext as well
155 as references were screened to identify additional PREMs. For all non-German PREMs
156 fulfilling the eligibility criteria additional searches were performed in the above-mentioned
157 databases to identify German translations. In addition, Google and google scholar were search
158 with the name of the PREM in combination with “German translation” or “cross-cultural
159 validation” to identify German translations. Only PREMs developed in German or for which a
160 German language translation existed were considered for further analyses.

161 **Data collection process**

162 Data was extracted on prespecified forms by two reviewers (AMi, CDH). The following data
 163 items were collected: author, year, journal, PREM acronym, country of origin, description of
 164 the PREM, method and timepoint of PREM collection, number of questions, PREM domains
 165 according to original article, presence of a German version, free text commentary.

166 **Assessment of psychometric properties and data synthesis**

167 Psychometric properties were used according to the COnsensus-based Standards for the
 168 selection of health Measurement INstruments (COSMIN) terminology [13]. Qualitative
 169 analysis of psychometric properties was done in according to current COSMIN guidelines
 170 [14,15] (www.cosmin.nl). The evaluation of content validity was done according to Terwee et
 171 al. [11]. We used the 16 PC dimensions described in S1 to assess content validity. The
 172 following psychometric properties were analysed: (1) content validity; (2) structural validity;
 173 (3) internal consistency; (4) measurement invariance / cross-cultural validity; (5) reliability;
 174 (6) measurement error; (7) criterion validity; (8) hypothesis testing for construct validity. Table
 175 1 shows the COMSIN assessment criteria used in this study. For overall evaluation “+” marks
 176 adequate, “-“ inadequate and “?” unclear psychometric properties. A detailed description of the
 177 methods can be found in [S5 Text](#).

178

179 **Table 1.** Criteria for overall assessment of psychometric properties according to COSMIN
 180 (Mokkink et al. 2018, COSMIN User Manuel).

181

Measurement property	Rating	Criteria
Structural validity	+	<p>CTT: CFA: CFI or TLI or comparable method >0.95 OR RMSEA <0.06 OR SRMR <0.082</p> <p>IRT/Rasch: No violation of unidimensionality³: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08</p>

		AND no violation of local independence: residual correlations among the items after controlling for the dominant factor < 0.20 OR $Q3's < 0.37$ AND no violation of monotonicity: adequate looking graphs OR item scalability > 0.30 AND adequate model fit: IRT: $\chi^2 > 0.01$ Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z standardized values > -2 and < 2
	?	CTT: Not all information for '+' reported IRT/Rasch: Model fit not reported
	-	Criteria for '+' not met
Internal consistency	+	At least low evidence for sufficient structural validity AND Cronbach's alpha(s) ≥ 0.70 for each unidimensional scale or subscale
	?	Criteria for "At least low evidence for sufficient structural validity" not met
	-	At least low evidence for sufficient structural validity ⁵ AND Cronbach's alpha(s) < 0.70 for each unidimensional scale or subscale
Measurement invariance (cross-cultural validity)	+	No important differences found between group factors (such as age, gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's $R^2 < 0.02$)
	?	No multiple group factor analysis OR DIF analysis performed
	-	Important differences between group factors OR DIF was found
Reliability	+	ICC or weighted Kappa ≥ 0.70
	?	ICC or weighted Kappa not reported
	-	ICC or weighted Kappa < 0.70
Measurement error	+	SDC or LoA $< MIC$
	?	MIC not defined
	-	SDC or LoA $> MIC$
Criterion validity	+	Correlation with gold standard ≥ 0.70 OR AUC ≥ 0.70
	?	Not all information for '+' reported
	-	Correlation with gold standard < 0.70 OR AUC < 0.70
Hypothesis testing for construct validity	+	The result is in accordance with the hypothesis
	?	No hypothesis defined (by the review team)
	-	The result is not in accordance with the hypothesis

182 "+=" sufficient, "-" = insufficient, "?" = indeterminate.

183 AUC: Area under the curve. CTT: classical test theory. CFA: confirmatory factor analysis, CFI: comparative fit index. DIF:
184 differential item functioning. ICC: interclass correlation coefficient. IRT: item response theory. MIC: minimal important
185 change. LoA: limits of agreement. RMSEA: Root Mean Square Error of Approximation. SDC: smallest detectable change.
186 SRMR: Square Root Mean Residuals. TLI: Tucker-Lewis index

187

188 **Results**

189 **Study selection and study characteristics**

190 By database searching 3,646 articles were identified. An additional 55 potential studies were
191 found by hand-search and 22 PREMs were added through the systematic review by Bull et al.
192 [8] (Fig 1). After screening abstracts and titles, 112 fulltext articles remained of which 51
193 articles fulfilled all eligibility criteria. Twelve of these 51 articles described German PREMs
194 [16–27] (Table 2). Details about the 39 non-German PREMs can be found in S6 Table. Of the
195 12 German-language PREMs, eight were generic [16,18–20,24–27] and four were cancer-care
196 specific PREMs [17,21–23]. No surgery-specific German-language PREM was identified. One
197 of the German-language PREMs (*Quality from the Patients' Perspective, QPP*) is available
198 only in its short-form (*short-form, QPPS*) [28]. The PREMs are either translations into German
199 [16–22] or were developed in German [23–26]. Study characteristics can be found in Table 2.

200

201 **Fig 1. PRISMA Flow-chart of included studies.**

202

203 **Table 2.** Study characteristics of German-language Patient-Reported Experience Measures (PREMs).

Author	Ref.	Year	Acronym	Country	Description	Method (timepoint of assessment)	Number of items	Domains	Comments	
Generic PREMs										
1	Oltedal S, Garratt A, Bjertnaes Ø, Bjørnsdottir M, Freil M, Sachs M.	Scandinavian Journal of Public Health, 2007; 35: 540–547	2007	NORPEQ	Norway	Further publications ins Skudal KE, Garratt AM, Eriksson B, Leinonen T, Simonsen J, Bjertnaes OA, BMJ Open. 2012;2(3):11.	paper-based (within 7 weeks after discharge)	8	<ol style="list-style-type: none"> Whether doctors were understandable Doctors' professional skills Nurses' professional skills Nursing care Whether doctors and nurses were interested in the patients' problems Information on tests 	German version used in Groene O, et al. Patient Experience Shows Little Relationship with Hospital Quality Management Strategies. PLoS One. 2015 Jul 7;10(7).
2	Jenkinson C, Coulter A, Bruster S.	Int J Qual Health Care. 2002;14(5):6.	2002	Picker PPE-15	Great Britain	“(1) to determine what factors are most likely to influence satisfaction with care and willingness to recommend hospital services to others and (2) to explore the extent to which satisfaction is a meaningful indicator of patients' experience of healthcare services.”	paper-based (within 1 month after discharge)	15	<ol style="list-style-type: none"> information and education coordination of care physical comfort emotional support respect for patient preferences involvement of family/friends continuity of care. 	
3	Noest S, Ludt S, Klingenberg A, et al.	Int J Qual Health Care. 2014;26(3):10.	2014	PEACS	Germany	“To develop and validate a generic questionnaire to evaluate experiences and reported outcomes in patients who receive treatment across a range of healthcare sectors”	paper-based (after discharge; exact timepoint not specified)	59 (total) 40 (short-form)	<ol style="list-style-type: none"> Preliminary care Shared decision-making at indication Patient education and information Nursing staff Accessible physicians Pain therapy Institutional treatment and transition Information at discharge follow-up Outcome 	
4	Hays RD, Shaul JA, Williams VS, et al.	Med Care. 1999;37(3 Suppl):MS22-31.	1999	HCAHPS CAHPS	USA	Further references in: 1.) Keller S, O'Malley A, Hays RD, Mathew RA, Zaslavsky AM, Hepner AM, et al. Methods used to streamline the CAHPS Hospital Survey. Health Serv Res. 2005;40:2057–77. 2.) Hargraves JL, Hays RD, Cleary PD. Psychometric properties of the Consumer Assessment of Health Plans Study (CAHPS) 2.0 adult core survey. Health Serv Res. 2003;38(6 Pt 1):1509-1527.	paper-based or telephone-interview	29	<ol style="list-style-type: none"> Global rating items <ol style="list-style-type: none"> Health plan Quality of care Personal doctor Specialist 10 domains of health plan performance <ol style="list-style-type: none"> Getting care you need Getting care without long wait Communication of doctors Doctors spending enough time with their patients Prevention Office staff Customer service 	German translation in Squires A, Bruyneel L, et al. Cross-cultural evaluation of the relevance of the HCAHPS survey in five European countries. Int J Qual Health Care. 2012 Oct;24(5):470-5. Further references in: 6.) Giordano L, Ellito M, Goldstein E, Lehrman E, Spencer P. Med Care Res Rev. 2009;67:27–37

- 3.) Hays RD, Martino S, Brown JA, et al. Evaluation of a care Med Care Res Rev. 2014;71(2):192-202
- 4.) Sofaer S, Crofton C, Goldstein E, Hoy E, Crabb J. Health Serv Res. 2005;40:2018–36
- 5.) Agency for Healthcare Research and Quality. HCAHPS Fact Sheet CAHPS Hospital Survey. USA; 2012

8. Reasonable paperwork
9. Finding a personal doctor
10. Referrals to specialists

- 7.) Levine RE, Fowler FJ, Brown JA. Health Serv Res. 2005;40:2037–56
- 8.) O'Malley AJ, Zaslavsky AM, Hays RD, Hepner KA, Keller S, Cleary PD. Health Serv Res. 2005;40:2078–94

5	Larsson BW, Larsson G.	J Clin Nurs. 2002;11:681–7.	QPP-S (Quality from the Patients' Perspective Shortened)	Sweden	short-form of the QPP	paper-based (at discharge)	24	<ol style="list-style-type: none"> 1. Medical-technical competence 2. Physical technical conditions 3. Identity-orientated approach 4. Socio-cultural atmosphere 	<p>German translation in</p> <ol style="list-style-type: none"> 1.) Singer S, Götze H, Möbius C, Witzigmann H, Kortmann RD, Lehmann A, Höckel M, Schwarz R, Hauss J. Langenbecks Arch Surg. 2009 Jul;394(4):723–31. doi: 10.1007/s00423-009-0489-5. 2.) Singer S, Danker H, Meixensberger J, Briest S, Dietz A, Kortmann RD, Stolzenburg JU, Kersting A, Roick J. J Cancer Res Clin Oncol. 2019 Nov;145(11):2845-2854.
6	Rosa Suñol, Oliver Groene, the DUQuE Project Consortium	http://www.duque.eu/index.php?page=participating-countries	DUQUE (Deepening our Understanding of Quality Improvement in Europe)	Czech Republic, France, Germany, Poland, Portugal, Spain, Turkey, Great Britain	<p>Set of patient-reported experience measures including</p> <ul style="list-style-type: none"> • a generic 6-item measure of patient experience (NORPEQ) • a 3-item measure of patient-perceived discharge preparation (Health Care Transition Measure) • perceived involvement in care (from Commonwealth Fund sicker patients survey) • hospital recommendation (from HCAHPS) 	paper-based	24	<ol style="list-style-type: none"> 1. Generic patient experience 2. Perceived patient involvement 3. Hospital recommendation Perceived continuity of care 	<p>The German version of the questionnaire has not yet been validated.</p> <p>Further references in:</p> <p>Groene O, Arah OA, Klazinga NS, Wagner C, Bartels PD, Kristensen S, Saillour F, Thompson A, Thompson CA, Pfaff H, DerSarkissian M, Sunol R. Patient Experience Shows Little Relationship with Hospital Quality Management Strategies. PLoS One. 2015 Jul 7;10(7):e0131805.</p>
7	Christoph Gehrlach, Thomas Altenhöner, David Schwappach	Der Patients' Experience Questionnaire. Patientenerfahrungen vergleichbar machen. Verlag Bertelsmann Stiftung. ISBN: 978-3-86793-021-5. 2009.	PEQ (German) (Patient Experience Questionnaire)	Switzerland, Germany	Developed in Switzerland and Germany	paper-based (14 days till 8 weeks after discharge. Two reminders after 14 and 28 days)	15	<ol style="list-style-type: none"> 1. Care by physicians (<i>Ärztliche Versorgung</i>) 2. Care by nurses (<i>Pflegerische Betreuung</i>) 3. Organisation and Service 4. Recommendation (<i>Weiterempfehlung</i>) 	There is a birth-specific version (PEQ-Geburt)
8	Tonio Schoenfelder, Joerg Klewer, Joachim Kugler	International Journal for Quality in Health Care, Volume 23, Issue 5, October 2011, Pages 503–509,	N/A	Germany	"Rates elements of satisfaction for cancer patients."	paper-based (after treatment)	28	<ol style="list-style-type: none"> 1. Global patient satisfaction 2. Medical aspects of care 3. Performance of service Patient expectations 	Application in surgery: Schoenfelder T, Klewer J, Kugler J. Factors associated with patient satisfaction in surgery: the role of patients' perceptions of received care, visit characteristics, and demographic

cancer care-specific PREMs

1	Brédart A, Ravazi D, Delvaux N, Goodman V, Farvacques C, Van Heer C.	A comprehensive assessment of satisfaction with care for cancer patients. Support Care Cancer. 1998;6(6):518–23	1998	EORTC IN-PATSAT32 (in-patient satisfaction with cancer care)	Europe	For in-hospital cancer patients independent of treatment discipline	paper-based (2 weeks after discharge)	32	<ol style="list-style-type: none"> 1. Technical quality 2. Interpersonal manner 3. Communication skills 4. Coordination 5. Waiting time 6. Continuity 7. Availability 8. Access 9. Physical environment 	To assess the <i>patient experience</i> of cancer patients treated in hospital.
2	Bokemeyer F, Lange-Drenth L, Jean-Pierre P, Schulz H, Bleich C.	Psychometric evaluation of the German version of the Patient Satisfaction with Cancer-related Care questionnaire. BMC Health Serv Res. 2020 Oct 27;20(1):983.	2020	PSCC-G (German Version of the Patient Satisfaction with Cancer Related Care)	Germany	Translation of the PSCC with additional questions from the PASQCOC questionnaire and the RESPERES-G questionnaire	paper-based	18	<ol style="list-style-type: none"> 1. Patient as a unique person 2. Essential characteristics of the clinician 3. Patient involvement in care 4. Emotional support 5. Clinician-patient communication 6. Family and friends 7. Patient empowerment 8. Patient information 9. Access to care 10. Integration of medical and non-medical care 11. Coordination of care 12. Teamwork 13. Clinician-patient relationship 14. Effective treatment 	<ul style="list-style-type: none"> - German translation - Original PSCC: Jean-Pierre P, Fiscella K, et al. Cancer. 2011;117(4):854–61. and Jean-Pierre P, Fiscella K, et al., Support Care Cancer. 2012;20(9):1949–56. - Additional questions from Recherché Evaluative sur la Performance de Réseau de Santé-German (RESPERES-G) and Patient Satisfaction and Quality in Oncological Care (PASQCOC)
3	Rudolph C, Petersen GS, Pritzkeleit R, Storm H, Katalinic A.	The acceptance and applicability of a patient-reported experience measurement tool in oncological care: a descriptive feasibility study in northern Germany. BMC Health Serv Res. 2019 Nov 1;19(1):786.	2019	Danish National Cancer Patient Questionnaire	Denmark	Translation of the Danish National Cancer Patient Questionnaire (Version 2017) and Feasibility testing in colorectal and breast cancer patients	paper-based (6 till 9 months after diagnosis)	99	<ol style="list-style-type: none"> 1. Diagnostic delay and satisfaction with diagnostic phase 2. Information & involvement 3. Continuity of care 4. Help & support 5. Discharge and overall treatment 	German translation. „...modifications of questions heavily related to the Danish health care system, especially referring to the diagnostic phase, are necessary “
4	Esser P, Sautier L, Sarkar S, Schilling G, Bokemeyer C, Koch U, Friedrich M, Defossez G, Mehnert-Theuerkauf A.	Development and preliminary psychometric investigation of the German Satisfaction with Comprehensive Cancer Care (SCCC) Questionnaire. Health Qual Life Outcomes. May 17;19(1):147.	2021	SCCC Satisfaction with Comprehensive Cancer Care	Germany	Assessment of patient satisfaction with „Comprehensive Cancer Care“	paper-based (in-hospital and outpatients. Timepoint of assessment not defined)	32	<ol style="list-style-type: none"> 1. Competence and human quality of physician (Competence) 2. Quality and quantity of information (Information) 3. If needed: access to psychosocial support services (Access) 4. Psychological support by medical staff (Support) 5. Global satisfaction with medical care 	Based on the French REPERES-60 questionnaire. (for breast cancer patients). Additional questions evaluating psychosocial support.

204

205

206

207

208

209 **Content validity**

210 Content validity was evaluated using the 16 dimensions of PC outlined in the introduction
211 (Table 3). None of the questionnaires covers all aspects of PC. An overview of the
212 methodological quality of studies analyzing the content validity according to COSMIN can be
213 found in Table 4. All 12 included PREMs exhibit sufficient relevance. Because of the lack of
214 content dimensions all questionnaires have deficits in *comprehensiveness* (Table 4).
215 Understandability of questionnaires is adequate in most cases. However, the methodological
216 quality of content validity studies varies widely from low (Schoenfelder, DUQUE) to high
217 (Picker, HCAHPS, PEQ, EORTC IN-PATSAT32, PSCC-G) (Table 4).

218

219 **Table 3.** Overview of content domains of German-language PREMs.

	PREM Domain	HCAHPS	PEACS	PPE-15	NORPEQ	QPP*	QPPS	DUQUE	PEQ (German)	Schoenfelder et al.	EORTC IN-PATSAT32	PSCC-G	Danish National CPQ	SCCC
1	patient as a unique person	-	-	-	-	+	+	-	+	-	+/-	+	+	+
2	biopsychosocial perspective	-	-	-	-	+/-	-	-	-	+	-	-	-	+
3	essential characteristics of the clinician	-	-	-	+/-	-	-	+/-	-	+	+	+	+	+
4	patient involvement in care	-	+	+	+	+	+/-	+	+	+	-	+/- ⁴	+	+
5	involvement of family and friends	-	+	+	-	+		+	-	-	-	+/- ⁴	-	-
6	physical support	+	+	+	+/-	+	+	-	-	-	+		+	+
7	emotional support	-	+	+	+/-	+/-	+/-	+	-	-	+	+/-	+	+
8	clinician-patient communication	+	+	+	+	+/-	+/-	+	-	+	+	+	+	+
9	patient empowerment	-	+	-	-	+/-	-	+	-	-	-	+	-	+
10	patient Information	+	+	-	+	+	+/-	+	+	+	+	+/- ⁴	+	+
11	access to care	-	+	-	-	-	-	-	+	-	+/-	+	-	+
12	integration of medical and non-medical care	-	+	-	-	-	-	-	-	-	-	+	-	+/-
13	coordination and continuity of care	+/-	+	+	-	-	-	+/-	+/-	+	-	+	+	+/-
14	teamwork and teambuilding	-	+	-	-	-	-	-	-	-	+/- (only nursing care)	+ ¹	-	-
15	clinician-patient relationship	+/-	+	-	+/-	+	+/-	+/-	+	-	+	+	+	+
16	effective treatment	-	+	-	+/-	+	+	-	+	+	+	+	-	+/-

220 **Table 4.** Rating of content validity of German PREMs according to COSMIN.

		Relevance					Comprehensiveness	Comprehensibility				Quality of Evidence*	Quality of methods*	Comment
		1	2	3	4	5	6	7	8	9	10			
		Are the included items relevant for the construct of interest ?	Are the included items relevant for the target population of interest ?	Are the included items relevant for the context of use of interest ?	Are the response options appropriate?	Is the recall period appropriate ?	Are no key concepts missing?	Are the PROM instructions understood by the population of interest as intended?	Are the PROM items and response options understood by the population of interest as intended ?	Are the PROM items appropriately worded?	Do the response options match the question ?			
generic PREMs														
1	NORPEQ	+	+	+	+	+	-	+	+	+	+	moderate	moderate	no cross-cultural validation in German
2	Picker PPE-15	+	+	+	+	+	+/-	+	+	+	+	moderate	high	unclear whether surgical patients were involved in design of the questionnaire
3	PEACS	+	+	+	+	?	+	+	+	+	+	high	moderate	
4	HCAHPS	+	+	+	+	+/-	-	+	+	+	+	moderate	high	
5	QPPS	+	+	+	+	+	-	?***	?***	?***	+	high**	high**	Short-version of the QPP. No surgical patients involved in the design of the questionnaire. No cross-cultural validation in German.
6	DUQUE	+	+	+	+	?	-	?***	?***	?***	+	low	low	
7	PEQ (German)	+	+	+	+	+	-	+	+	+	+	high	high	
8	Schoenfelder et al.	+	+	+	+	?	-	+	+	+	+	low	low	

Cancer-care specific PREM														
1	EORTC IN-PATSAT32	+	+	+	+	+	+/-	+	+	+	+	high	high	
2	PSCC	+	+	+	+	+	+	+	+/-	+	+	moderate	high	
3	Danish National Cancer Patient Questionnaire	+	+	+/- ****	+	+/-	+/-	+/-	+	+	+	low*** (German translation)	low*** (German translation)	
4	PASQOC	+	+	+	+	?	+/-*****	+	+	+/-	+	moderate	moderate	
5	SCCC	+	+	+	+	?	+/-	+/-	+	+/-	+	moderate	moderate	Based on the REPERES-60, which was designed for breast cancer patients.

* high, moderate, low, very low

** rating is based on the long-version (QPP) rather than QPPS

*** lack of validation study in German

**** authors recommend to omit several items in the German version

*****designed for the outpatient setting

221

222

223

224 **Structural validity**

225 “Structural validity refers to the degree to which the scores of a PROM (or PREM) are an
226 adequate reflection of the dimensionality of the construct to be measured” [11]. A summary
227 table of the findings on structural validity can be found in Table 5. Not all necessary data on
228 confirmatory factor analyses according to current COSMIN guidelines were available to rate
229 the NORPEQ or the questionnaire by Schoenfelder et al. (Table 5). The PPE-15 received an
230 insufficient rating for structural validity due to the inadequate design of the underlying studies
231 [16]. Similarly, the SCC showed insufficient structural validity. No data on structural validity
232 could be obtained for the German PEQ neither in the original publication nor in additional
233 studies [25]. The same was true for the Danish National Cancer Questionnaire. PEACS and
234 HCAHPS received a sufficient rating (Table 5). For the EORTC INPAT-SAT32 there is a
235 metanalysis summarizing the data on structural validity [29]. There is a detailed analysis of
236 structural validity for the PSCC-G [22], which shows mixed results.

237

238 **Table 5.** Overview of the analyses results on structural validity of German-language PREMs.

		Reference	1 Is the PREM based on a reflective or formative model?	2 Does the study concern unidimensionality?	3a For CTT: Was exploratory or confirmatory factor analysis performed?	3b For IRT/Rasch: does the chosen model fit to the research question?	4 Was the sample size included in the analysis adequate?	5 Were there any other important flaws in the design or statistical methods of the study?	6 Overall rating on structural validity
generic PREMs									
1	NORPEQ	Oltedal S, et al. Scandinavian Journal of Public Health, 2007; 35: 540–547	reflective	no	very good (confirmatory and explorative)	N/A	very good (n=244 participants for 8 questions)	very good	?*
2	Picker PPE-15	Jenkinson C, Coulter A, Bruster S. Int J Qual Health Care. 2002;14(5):6.	reflective	yes	inadequate	N/A	vey good	inadequate	-
3	PEACS	Noest S, Ludt S, Klingenberg A, et al., Int J Qual Health Care. 2014;26(3):10.	reflective	yes	very good (confirmatory and explorative)	N/A	very good (n=492 participants for 59 questions)	very good	+
4	HCAHPS	Keller S, et al. Health Serv Res. 2005;40:2057–77.	reflective	yes	very good (confirmatory and explorative)	N/A	very good (n=19,720 participants)	very good	+

5	QPPS	No data available on structural validity for the QPPS	reflective	0	0	N/A	0	0	?
	QPP	Larsson G, Larsson BW, Munck IME. Scand J Caring Sci. 1998;12:111–8.;	reflective	yes	very good (confirmatory)	N/A	very good (n=611 participants)	very good	+
6	DUQUE	zum DUQUE sind keine Studien zur Strukturvalidität vorhanden	formative	0	0	N/A	0	0	?
7	PEQ (German)	zum PEQ (German) sind keine Studien zur Strukturvalidität vorhanden	reflective	0	0	N/A	0	0	?
8	Schoenfelder et al.	Tonio Schoenfelder, et al. . International J. Quality in Health Care, Volume 23, Issue 5, October 2011, Pages 503–509	reflective	yes	adequate (explorative)	N/A	very good (n=5,774 participants)	very goodt	?*
cancer-care specific PREMs									
1	EORTC IN-PATSAT32	Neijenhuijs KI, et al. Support Care Cancer. 2018 Aug;26(8):2551-2560.	reflective	yes	very good (confirmatory)	N/A	very good	inadequate	-

2	PSCC-G	Bokemeyer F, Lange-Drenth L, Jean-Pierre P, Schulz H, Bleich C. BMC Health Serv Res. 2020 Oct 27;20(1):983	reflective	yes	very good (confirmatory and explorative)	N/A	very good (n=394 participants, 18 questions)	very good	+/-**
3	Danish National Cancer Patient Questionnaire	Rudolph C, Petersen GS, Pritzkeleit R, Storm H, Katalinic A. BMC Health Serv Res. 2019 Nov 1;19(1):786 Danish Cancer Society.	reflective	0	0	N/A	0	0	?
		Kræftpatienters behov og oplevelser med sundhedsvæsenet under udredning og behandlingKræftens Bekæmpelse, 2017. ISBN: 978-87-7064-367-2	reflective	?	0	N/A	very good	0	
4	SCCC	Esser P, Sautier L, Sarkar S, Schilling G, Bokemeyer C, Koch U, Friedrich M, Defossez G, Mehnert-Theuerkauf A. Health Qual Life Outcomes. 2021 May 17;19(1):147	reflective	yes	adequate (explorative)	N/A	very good (n=333 participants, 32 questions)	doubtful	-

“+” = sufficient, ”-“ = insufficient, “?” = unclear

* not all necessary information for a "+" rating available.

** all data available, but mixed results

0 = not reported

N/A = not applicable

239

240

241 **Internal consistency**

242 Table 6 shows the results of internal consistency studies of German-language PREMs.
243 Cronbach's Alpha for the NORPEQ (containing only one scale) is 0.85 [20]. A single
244 validation study is available for the PPE-15 [16]. It is unclear whether the consistency statistics
245 contained in this study are calculations of Cronbach's Alpha or not. Furthermore, the
246 consistency analyses are not available for the various PPE-15 subscales, which are partly
247 dichotomous, partly continuous. Consequently, the PPE-15 received an "insufficient" rating
248 for study design. Similarly, internal consistency analyses (Cronbach's Alpha) of the PEQ refer
249 to the entire questionnaire and not to the various subscales [25]. There are no internal
250 consistency studies for the PEACS questionnaire. Keller et al. reported data on internal
251 consistency for all 7 subscales of the HCAHPS [19]. With exception of the subscale "*Medicine*
252 *Communication*" (Cronbach's Alpha 0.66) all HCAHPS subscales show a Cronbach's Alpha
253 ≥ 0.70 . Analysis of internal consistency is available for all subscales of the QPPS in the study
254 by Larsson et al. [18]. Results are mixed as some subscales exhibit Cronbach's alpha ≥ 0.70
255 and some < 0.70 . No data on internal consistency could be found for the DUQUE and the
256 *Danish National Cancer Patient Questionnaire*. For the PREM by Schoenfelder et al. internal
257 consistency results are available for both subscales of the questionnaire (*received service* and
258 *medical aspects*). Both of which have a Cronbach's Alpha ≥ 0.70 .
259 The internal consistency of the EORTC IN-PATSAT32 has been analyzed in a recent
260 metaanalysis [29]. According to this metaanalysis, 5 studies are available on the internal
261 consistency of the IN-PATSAT32, of which 3, however, are of poor methodological quality.
262 Analyses are available for all subscales of the questionnaire, with Cronbach's Alpha ≥ 0.70
263 across all scales, except for the subscale *hospital access* (Cronbach's Alpha < 0.70). The overall
264 Cronbach's Alpha for the entire PSCC-G is reported to be 0.92 [22]. As the German version of
265 the PSCC (PSCC-G) consists of the German translation of the English PSCC [30] as well as of

266 translations of parts of the French REPERES (subscale „Information“) [31] and German
267 PASQOC [32] (subscales *family and friends*, *shared-decision making* and *nursing staff*), the
268 original questionnaires can be analyzed for internal consistency as well. For all subscales
269 Cronbach’s Alpha is reported to be ≥ 0.70 .

270

271

272 **Table 6.** Overview of internal consistency studies of German-language PREMs. Rating according to COSMIN guidelines.

			1	2	3	4	
	Reference		Does the scale consist of effect indicators, i.e., is it based on a reflective model	Was an internal consistency statistic calculated for each unidimensional scale or subscale separately?	For continuous scores: Was Cronbach's alpha or omega calculated?	For dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	Overall rating
generic PREMs							
1	NORPEQ	Oltedal S, et al. Scandinavian Journal of Public Health, 2007; 35: 540–547	reflective	very good (contains only one scale)	very good (calculation of Cronbach's Alpha)	N/A	+
2	Picker PPE-15	Jenkinson C, et al. Int J Qual Health Care. 2002;14(5):6.	reflective	insufficient (see text)	very good (calculation of Cronbach's Alpha)	insufficient	?
3	PEACS	Noest S, et al., Int J Qual Health Care. 2014;26(3):10.	reflective	0	0	0	?
4	HCAHPS	Keller S, et al. Health Serv Res. 2005;40:2057–77.	reflective	very good (calculation for each subscale)	very good (calculation of Cronbach's Alpha)	N/A	+/-
5	QPPS	Larsson BW, Larsson G. J Clin Nurs. 2002;11:681–7.	reflective	very good (calculation for each subscale)	very good (calculation of Cronbach's Alpha)	N/A	+/-**
6	DUQUE	No studies could be identified	formative	0	0	0	?
7	PEQ (German)	No studies could be identified	reflective	insufficient (see text)	very good (calculation of Cronbach's Alpha)	N/A	?
8	Schoenfelder et al.	Tonio Schoenfelder, et al. Journal for Quality in Health Care, Volume 23, Issue 5, October 2011, Pages 503–509	reflective	very good (calculation for each subscale)	very good (calculation of Cronbach's Alpha)	N/A	+
cancer-care specific PREMs							
1	EORTC IN-PATSAT32	Neijenhuijs KI, et al. Support Care Cancer. 2018 Aug;26(8):2551-2560.	reflective	very good (calculation for each subscale)	very good (calculation of Cronbach's Alpha)	N/A	+/-

	PSCC-G	Bokemeyer F, et al. BMC Health Serv Res. 2020 Oct 27;20(1):983	reflective	sufficient	very good (calculation of Cronbach's Alpha)	N/A	+
	PSCC (engl. Original version)	Jean-Pierre P, et al, Group PNRP. Cancer. 2011;117(4):854-61	reflective	sufficient	very good (calculation of Cronbach's Alpha)	N/A	+
2	RESPERES (subscale "Information"; part of the German PSCC-G)	Defossez G, et al. Health Qual Life Outcomes. 2021 May 17;19(1):147	reflective	sufficient	very good (calculation of Cronbach's Alpha)	N/A	+
	PASQOC (subscale "family and friends", "shared-decision making" & "nursing" part of the German PSCC-G)	Kleeberg UR, et al. M. Support Care Cancer. 2008;16(8):947-54.	reflective	sufficient	very good (calculation of Cronbach's Alpha)	N/A	+
		Rudolph C, et al. BMC Health Serv Res. 2019 Nov 1;19(1):786	reflective	0	0	0	?
3	Danish National Cancer Patient Questionnaire	Danish Cancer Society. Kræftpatienters behov og oplevelser med sundhedsvæsenet under udredning og behandling. Kræftens Bekæmpelse, 2017. ISBN: 978-87-7064-367-2	reflective	0	0	0	?
4	SCCC	Esser P, et al. Health Qual Life Outcomes. 2021 May 17;19(1):147	reflective	insufficient	very good (calculation of Cronbach's Alpha)	N/A	+

“+” = sufficient, ”-“ = insufficient, “?” = unclear

* not all information available for "+" rating.

** all data available, but mixed results (see text)

0 = not reported

N/A = not applicable

275 **Measurement invariance / Cross-cultural validity**

276 “Cross-cultural validity refers to the degree to which the performance of the items on a
277 translated or culturally adapted instrument are an adequate reflection of the performance of the
278 items of the original version of the instrument” [11]. It is particularly important for PREM
279 translations into German. Results are depicted in Table 7. No data on cross-cultural validity
280 could be found for the DUQUE questionnaire. A cross-cultural validation study for the
281 NORPEQ is available is available for several Scandinavian countries, but not for German. The
282 PPE-15 underwent validation in 5 countries including German speaking countries (Switzerland
283 and Germany) [16]. The patient cohorts were comparable in respect to sex and age, but not in
284 respect to indications (elective vs. emergency). No further details about the patient cohorts
285 were collected. Furthermore, no adequate study could be identified that compares the English
286 original questionnaire with the German translation [16]. Such a study exists for the Spanish
287 language [33]. This study shows that an expansion of the questionnaire from 15 to 33 questions
288 (PPE-33) is necessary to preserve psychometric properties [33]. There is a cross-cultural
289 validation study for the HCAHPS PREM [34]. In this study it is unclear whether patient cohorts
290 were comparable. Furthermore, only 10 German-speaking patients were included [34]. In
291 addition, the statistical methods used are considered to be inadequate according to current
292 COSMIN guidelines. The QPPS has been used in two German studies without prior validation
293 of the German translation [28,35]. Validation of the EORTC IN-PATSAT32 occurred in an
294 international study in which 34 German-speaking patients were included [36]. However, cross-
295 cultural validity was not analyzed in this study. An addition, the EORTC IN-PATSAT32
296 metaanalysis by Neijenhuijs et al. found no studies on cross-cultural validation of the
297 questionnaire in German [29]. Similarly, no cross-cultural validation in German could be found
298 for the Danish *National Cancer Questionnaire* [37]. Rudolph et al. use the Danish *National*
299 *Cancer Questionnaire* in their study but without previous validation [21]. The PSCC-G has

300 been validated in German in a study by Bokemeyer et al. with a sufficient number of patients
301 and adequate statistical methods [22]. The remaining PREMs (PEACS, PEQ, Schoenfelder et
302 al, SCCC) have been developed in German including patients from different medical
303 specialities.

304

305

306

307 **Table 7.** Results of cross-cultural validity analyses of PREMs into German. Rating according to COSMIN guidelines.

308

		Reference	Were the samples similar for relevant characteristics except for the group variable?	Was an appropriate approach used to analyse the data?	Was the sample size included in the analysis adequate?	Overall rating
generic PREMs						
1	NORPEQ	Olstedal S, et al. Scandinavian Journal of Public Health, 2007; 35: 540–547	0	0	0	?
2	Picker PPE-15	Jenkinson C, Coulter A, Bruster S. Int J Qual Health Care. 2002;14(5):6.	doubtful	inadequate	adequate	?
3	PEACS	Noest S, Ludt S, Klingenberg A, et al., Int J Qual Health Care. 2014;26(3):10.	N/A*	N/A*	N/A*	N/A*
4	HCAHPS	Squire et al. International Journal for Quality in Health Care 2012; Volume 24, Number 5: pp. 470–475	doubtful	doubtful	inadequate	?
5	QPPS	Singer S, et al. Langenbecks Arch Surg. 2009, 394:723–731	0	0	0	?
		Singer S, Danker H, et al. J Cancer Res Clin Oncol. 2019	0	0	0	?
6	DUQUE	no studies on measurement invariance / cross-cultural validity could be found	0	0	0	?
7	PEQ (German)	Christoph Gehrlach, et al. Verlag Bertelsmann Stiftung. ISBN: 978-3-86793-021-5. 2009.	N/A*	N/A*	N/A*	N/A*
8	Schoenfelder et al.	Tonio Schoenfelder, Joerg Klewer, Joachim Kugle. International Journal for Quality in Health Care, Volume 23, Issue 5, October 2011, Pages 503–509	N/A*	N/A*	N/A*	N/A*
cancer-care specific PREMs						
1	EORTC IN-PATSAT32	Neijenhuijs KI, Jansen F, et al. Support Care Cancer. 2018 Aug;26(8):2551-2560.	0	0	inadequate	?

		A. Brédart, A. Bottomley, Jet al. European Journal of Cancer 41 (2005) 2120–2131	0	0	inadequate	N/A
2	PSCC-G	Bokemeyer F, Lange-Drenth L, Jean-Pierre P, Schulz H, Bleich C. BMC Health Serv Res. 2020 Oct 27;20(1):983	adequate	very good	very good	+
3	Danish National Cancer Patient Questionnaire	Rudolph C, et al. BMC Health Serv Res. 2019 Nov 1;19(1):786	0	0	0	?
		Danish Cancer Society. Kræftpatienters behov og oplevelser med sundhedsvæsenet under udredning og behandling Kræftens Bekæmpelse, 2017. ISBN: 978-87-7064-367-2	0	0	0	?
4	SCCC	Esser P, et al. Health Qual Life Outcomes. 2021 May 17;19(1):147	N/A*	N/A*	N/A*	N/A*

309
310
311
312

“+” = sufficient, “-” = insufficient, “?” = unclear
N/A: not applicable as the PREM has been developed in German.
0 : not reported

313

314 **Reliability**

315 An overview of the results of the reliability analyses can be found in table 8. A study measured
316 test-retest reliability of the NORPEQ [20] for which 68 of 244 patients were resent the
317 questionnaire within 5-6 days . The intraclass correlation index (ICC) was between 0.45 (*nurse*
318 *professional skills*) and 0,83 (*doctors understandable*). Four of the eight subscales exhibited
319 an ICC <0.70. The test-retest reliability for the total score was 0.88 [20]. Consequently, the
320 overall reliability rating for the NORPEQ was +/- . Noest et al. analyzed reliability for the
321 PEACS questionnaire [24]. Test-retest reliability was measured via weighted kappa. With
322 exception of the subscale institutional treatment and transition (weighted kappa 0,671), the
323 weighted kappa was ≥ 0.70 for all subscales. Thus, overall rating was + . The study by Keller et
324 al. analyzed hospital-level reliability [19]. No data on test-retest reliability could be identified
325 for the HCAHPS. Hospital-level reliability assumes that recurrent measurements (retesting) of
326 patients in the same hospital should be more similar than recurrent measurements in another
327 hospital. In the study by Keller et al. 300 patients from different hospitals were asked to fill out
328 the HCAHPS. For six of the seven subscales the ICC value was ≥ 0.70 . The only exception was
329 *medicine communication* (ICC 0,66).

330 The overall rating for the HCAHPS was + . A similar hospital-level reliability analysis can be
331 found for the PEQ in the study by Gehrlach et al [25]. Logistic regression analyses were used
332 to find out whether the questionnaire can distinguish between patient cohorts from different
333 hospitals. However, no specific results are reported except for "...none of the instruments
334 showed significant results". Further reliability data for the PEQ could not be identified. No
335 reliability study could be identified for the PPE-15, QPPS or the QPP, DUQUE, PSCC-G, the
336 Danish National Cancer Patient Questionnaire or the SCCC.

337 For the cancer-care specific PREM EORTC IN-PATSAT32 two studies investigate test-retest
338 reliability [38,39]. Appreciation of reliability in these two studies has already been done by

339 Neijenhuijs et al. [29]. ICC was ≥ 0.70 for all subscales of the EORTC IN-PATSAT32 in the
340 study by Pishkuhi et al. In the study by Obtel et al. all scales except *doctor availability*
341 (correlation coefficient 0.64) and *overall satisfaction* (correlation coefficient 0.67) showed a
342 correlation coefficient ≥ 0.70 . Consequently, overall reliability rating was +. However, both
343 studies showed methodological weaknesses as the time interval between test and retest was too
344 short (30 minutes) [39] or it was unclear, which type of correlation coefficient has been used
345 [38].
346
347

348 **Table 8.** Results of reliability analyses of German-language PREMs. Rating according to COSMIN guidelines.

		Reference	1 Were patients stable in the interim period on the construct to be measured?	2 Was the time interval appropriate?	3 Were the test conditions similar for the measurements?	4 For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	5 For dichotomous/nominal/ordinal scores: Was kappa calculated?	6 For ordinal scores: Was a weighted kappa calculated?	7 For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	8 Were there any other important flaws in the design or statistical methods of the study?	Overall rating
generic PREMs											
1	NORPEQ	Oltedal S, et al. Scandinavian Journal of Public Health, 2007; 35: 540–547	very good	very good (5-6 days)	very good	adequate	N/A	N/A	N/A	very good	+/-
2	Picker PPE-15	Jenkinson C, Coulter A, Bruster S. Int J Qual Health Care. 2002;14(5):6.	0	0	0	0	N/A	N/A	N/A	0	?
3	PEACS	Noest S, et al., Int J Qual Health Care. 2014;26(3):10.	very good	very good (3 weeks)	very good	N/A	very good	very good	N/A	very good	+
4	HCAHPS	Keller S, et al. Health Serv Res. 2005;40:2057–77.	doubtful	N/A*	adequate	adequate	N/A	N/A	N/A	very good	+
5	QPPS	No reliability studies for the QPPS or QPP could be identified	0	0	0	0	N/A	N/A	N/A	0	?

6	DUQUE	No reliability studies for the DUQUE could be identified	0	0	0	0	N/A	N/A	N/A	0	?
7	PEQ-G	Christoph Gehrlach, Thomas Altenhöner, David Schwappach. Verlag Bertelsmann Stiftung. ISBN: 978-3-86793-021-5. 2009.	doubtful	N/A*	doubtful	inadequate	N/A	N/A	N/A	?	-
8	Schoenfelder et al.	Tonio Schoenfelder, et al. Int.J for Quality in Health Care, Volume 23, 5, 2011, 503–509	0	0	0	0	N/A	N/A	N/A	0	?

cancer-care specific PREMs

1	EORTC IN-PATSAT32	Obtel, M. et al. Asian Pac J Cancer Prev 18, 1403–1409 (2017).	very good	inadequate	very good	adequate	N/A	N/A	N/A	doubtful	+
		Pishkuhi, M. A., et al. Asian Pac J Cancer Prev 15, 10121–10128 (2014).	very good	adequate	very good	doubtful	N/A	N/A	N/A	doubtful	
2	PSCC-G	No reliability studies for the PSCC-G could be identified	0	0	0	0	N/A	N/A	N/A	0	?
	PSCC (engl. original)	No reliability studies for the PSCC could be identified	0	0	0	0	N/A	N/A	N/A	0	?

3	Danish National Cancer Patient Questionnaire	No reliability studies for the could be identified	0	0	0	0	N/A	N/A	N/A	0	?
---	--	--	---	---	---	---	-----	-----	-----	---	---

4	SCCC	No reliability studies for the SCCC could be identified	0	0	0	0	N/A	N/A	N/A	0	?
---	------	---	---	---	---	---	-----	-----	-----	---	---

349 “+” = sufficient, ”-“ = insufficient, “?” = unclear

350 N/A: not applicable

351 0 : not reported

352 * for the HCAHPS and the PEG (German) studies measured “hospital level realibility”, no data on test-retest, inter-rater or intra-Rater reliability

353

354 **Analysis of measurement error**

355 An overview of the results for the analysis of measurement error can be found in Table 9.
356 Measurement error could not be analyzed as *no minimal important change* has been defined
357 for any of the German PREMs so far. Therefore, all PREMs received “?” rating. Only for
358 HCAHPS there has been a calculation of the standard error of Measurement (SEM). For the
359 EORTC IN-PATSAT32 the SEM and SDC can be calculated from the studies by Obtel et al.
360 [39] and Pishkuhi et al. [38] as has been shown by Neijenhuijs et al. [29]. However, as no
361 *minimal important change* has been defined for the EORTC IN-PATSAT32 an overall rating
362 of the measurement error is not possible.

363

364

365

366 **Table 9.** Results of analyses for measurement error of German-language PREMs. Rating according to COSMIN guidelines.

		Reference	1 Were patients stable in the interim period on the construct to be measured?	2 Was the time interval appropriate?	3 Were the test conditions similar for the measurements?	4 For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	5 For dichotomous/ nominal/ ordinal scores: Was the percentage (positive and negative) agreement calculated?	6 Overall rating
generic PREMs								
1	NORPEQ	Oltedal S, et al. Scandinavian Journal of Public Health, 2007; 35: 540–547	very good	very good (5-6 days)	very good	0	0	?
2	Picker PPE-15	Jenkinson C, Coulter A, Bruster S. Int J Qual Health Care. 2002;14(5):6.	0	0	0	0	0	?
3	PEACS	Noest S, et al., Int J Qual Health Care. 2014;26(3):10.	very good	very good (3 weeks)	very good	0	0	?
4	HCAHPS	Keller S, et al. Health Serv Res. 2005;40:2057–77.	doubtful	N/A*	adequate	very good	N/A	?
5	QPPS	No studies on measurement error were identified	0	0	0	0	0	?
6	DUQUE	No studies on measurement error were identified	0	0	0	0	0	?
7	PEQ (German)	Christoph Gehrlach, et al. Verlag Bertelsmann Stiftung. ISBN: 978-3-86793-021-5. 2009.	doubtful	N/A*	doubtful	0	0	?
8	Schoenfelder et al.	No studies on measurement error were identified	0	0	0	0	0	?
cancer-care specific PREMs								

1	EORTC IN-PATSAT32	Obtel, M. et al. Asian Pac J Cancer Prev 18, 1403–1409 (2017).	very good	inadequate	very good	adequate	N/A	?
		Pishkuhi, M. A., et al. Asian Pac J Cancer Prev 15, 10121–10128 (2014).	very good	adequate	very good	adequate	N/A	
2	PSCC-G	No studies on measurement error were identified	0	0	0	0	0	?
3	Danish National Cancer Patient Questionnaire	No studies on measurement error were identified	0	0	0	0	0	?
4	SCCC	No studies on measurement error were identified	0	0	0	0	0	?

367

368 “+” = sufficient, ” –“ = insufficient, “?” = unclear

369 N/A: not applicable

370 0 : not reported

371

372

373

374 **Analysis of criterion validity**

375 As no gold standard for the measurement of patient-centeredness has yet been defined, overall
376 criterion validity cannot be analyzed for PREMs. However, for some PREM subscales, gold
377 standards for measurement are available. Consequently, criterion validity for these subscales
378 can be analyzed. No data on criterion validity was found for NORPEQ, PPE-15, HCAHPS,
379 QPPS (or QPP), DUQUE, Schoenfelder et al., EORTC IN-PATSAT-32, PSCC, SCC and the
380 Danish National Cancer Patient Questionnaire.

381 The PEACS subscales *shared decision making* and *information at discharge* were compared
382 to the two validated and established German questionnaires *Shared Decision Making*
383 *Questionnaire* (SDM-Q-9) [40] and the *Care-Transition-Measurement Questionnaire* (CTM)
384 [41], respectively. Both subscales showed a very high (SDM-Q-9; $r = 0.814$, $p < 0.001$) or high
385 correlation (CTM-3; $r = 0.511$, $p < 0.001$) with the respective gold standard [24].

386 The PEQ was correlated to the Cologne Patient Questionnaire (*Kölner Patientenfragebogen*,
387 *KPF*). For the PEQ subscales “physicians” and “nursing” there was a high correlation >0.70
388 with the KPF. Only a weak correlation was found for the subscale “management” (between -
389 0,28 and -0,46) [25].

390

391 **Analysis of hypothesis testing for construct validity**

392 Hypothesis testing for construct validity describes the degree to which a PREM results is
393 consistent with an *a priori* hypothesis. The hypothesis to be tested can either be a comparison
394 of PREM results between two clinically defined patient groups (*known groups validity*) or
395 PREM subscales can be compared to another known measurement tool (*convergent validity*).
396 No data on hypothesis testing for construct validity could be found for: PEACS, DUQUE,
397 PEQ-G, *Danish National Cancer Patient Questionnaire*. Results for all other German PREMs

398 can be found in Table 10. Most hypotheses are tested positive, i.e., results confirm the a priori
399 formulated hypothesis. However, all PREMs, except for the PSCC-G, also show negative
400 results. For the PSCC-G only positive test results could be found.

401

402

403 **Table 10.** Results of hypothesis testing for construct validity of German-language PREMs. Rating according to COSMIN guidelines.

	Reference	Hypothesis	Convergent validity			Known-groups validity		Result	Overall rating	
			Is it clear what the comparator instrument measures?	Were the measurement properties of the comparator instrument(s) sufficient?	Were design and statistical methods adequate for the hypotheses to be tested?	Was an adequate description provided of important characteristics of the subgroups?	Were design and statistical methods adequate for the hypotheses to be tested?			
generic PREMs										
1	NORPEQ	Oltedal S, et al. Scandinavian Journal of Public Health, 2007; 35: 540–547	A higher general satisfaction correlates positively with NORPEQ results	very good	very good	adequate	0	0	There is a correlation between NORPEQ scores and general satisfaction	+
		Oltedal S, et al. Scandinavian Journal of Public Health, 2007; 35: 540–547	Errors in care result in lower NORPEQ scores	very good	very good	adequate	0	0	There is a weak correlation between errors in care and NORPEQ results	+
		Oltedal S, et al. Scandinavian Journal of Public Health, 2007; 35: 540–547	Patients whose expectations have been met have higher NORPEQ scores	very good	very good	adequate	0	0	Results show the expected correlation	+
		Oltedal S, et al. Scandinavian Journal of Public Health, 2007; 35: 540–547	Patients with a higher general health status have higher NORPEQ scores	very good	very good	adequate	0	0	Results show the expected correlation	+
		Oltedal S, et al. Scandinavian Journal of Public Health, 2007; 35: 540–547	Age correlates positively with NORPEQ scores	0	0	0	very good	adequate	Results show the expected correlation	+
		Boge RM, Haugen AS, Nilsen RM, Bruvik F, Harthug S. BMJ Open Qual. 2019 Dec 16;8(4):e000728.	Patients that receive a discharge interview have higher NORPEQ scores	0	0	0	very good	inadequate	Results show a statistically significant difference. The correlation, however, is unclear	?

1	NORPEQ	Groene O, et al. PLoS One. 2015 Jul 7;10(7):e0131805.	Hospital quality initiatives increase the patient experience measured via the NORPEQ	0	0	0	very good	adequate	The expected correlation could not be found	-
		Leonardsen AL, et al. LP. BMC Health Serv Res. 2017 Sep 29;17(1):685.	Results of NORPEQ items correlate with comparable items of the PPE-15	very good	very good	very good	0	0	Results show a weak correlation between NORPEQ and PPE-15 items	+/-
2	Picker PPE-15	Harrison R, et al. Int J Qual Health Care. 2018 Jun 1;30(5):358-365.	Patients with an adverse event have higher PPE-15 scores	0	0	0	very good	adequate	Results show the expected correlation	+
		Leonardsen AL, et al. LP. BMC Health Serv Res. 2017 Sep 29;17(1):685.	Results of NORPEQ items correlate with comparable items of the PPE-15	very good	very good	very good	0	0	Results show a weak correlation between NORPEQ and PPE-15 items	+/-
		Wong EL, Leung MC, Cheung AW, Yam CH, Yeoh EK, Griffiths S. Int J Qual Health Care. 2011 Aug;23(4):390-6.	Patients in private hospitals in Hong-Kong have a better patient experience than patients in public hospitals	0	0	0	very good	very good	Results do not show higher PPE-15 scores in private hospitals	-
		Ashton F, Hamid K, Sulieman S, Eardley W, Baker P. Injury. 2017 Apr;48(4):960-965.	Patients whose surgery was postponed have lower PPE-15 scores	0	0	0	very good	inadequate	Results confirm the hypothesis, but no correlations were measured	?
		Ashton F, Hamid K, Sulieman S, Eardley W, Baker P. Injury. 2017 Apr;48(4):960-965.	Patients waiting >3 days for their operation have a lower score than patients waiting <3days	0	0	0	very good	inadequate	Results confirm the hypothesis, but no correlations were measured	?

2	Picker PPE-15	Ashton F, Hamid K, Suliman S, Eardley W, Baker P. <i>Injury</i> . 2017 Apr;48(4):960-965.	Patients undergoing arthroscopy have a higher score than patients undergoing ankle fracture surgery	0	0	0	very good	inadequate	Results confirm the hypothesis, but no correlations were measured	?
		Murrells T, Robert G, Adams M, Morrow E, Maben J. <i>BMJ Open</i> . 2013 Jan 30;3(1):e002211.	Results of the PPE-15 subscales correlate negatively with results of the Patient Evaluation of Emotional Care during Hospitalisation' (PEECH) subscales	very good	doubtful	adequate	0	0	Results show the expected negative correlation	+
		Andres EB, Song W, Song W, Johnston JM. <i>BMC Health Serv Res</i> . 2019 Sep 3;19(1):623.	Accredited hospitals have higher scores than non-accredited hospitals	0	0	0	very good	adequate	Results show the expected correlation	+
		Wolf A, Olsson LE, Taft C, Swedberg K, Ekman I. <i>BMC Nurs</i> . 2012 Jun 14;11:8.	Vulnerable patients have lower scores than non-vulnerable patients	0	0	0	very good	doubtful	Results show the expected correlation, however, groups differed in important aspects	+
3	PEACS	No studies on hypothesis testing for construct validity could be identified	0	0	0	0	0	0	0	?
4	HCAHPS	Keller S, O'Malley A, et al. <i>Bull Hosp Jt Dis</i> (2013). 2019 Dec;77(4):263-268.	Patients undergoing day-clinic surgery for hip-TEP have better scores than in-hospital patients undergoing the same surgery	0	0	0	very good	inadequate	Results show the expected correlation	+
		Huppertz JW, Smith R. <i>J Healthc Manag</i> . 2014 Jan-Feb;59(1):31-47.	positive and negative handwritten comments on hospital feedback questionnaires correlate with positive or negative HCHAPS results respectively	0	0	0	very good	inadequate	Results show the expected correlation	+

4	HCAHPS	Velez VJ, Kaw R, Hu B, Frankel RM, Windover AK, Bokar D, Rish JM, Rothberg MB. J Hosp Med. 2017 Jun;12(6):421-427.	The HCAHPS subscale “doctors communication” correlates with results of the Four Habits Coding Scheme (4HCS)	very good	very good	very good	0	0	Results did not show the expected correlation	-
		Huppertz JW, Otto P. Health Care Manage Rev. 2018 Oct/Dec;43(4):359-367.	HCAHPS scores correlate with social media comments	0	0	0	doubtful	adequate	Results show the expected correlation	+
		Bobrovitz N, et al. J Trauma Acute Care Surg. 2016 Jan;80(1):111-8.	For trauma patients HCAHPS scores correlate with results of the Quality of Trauma Care Patient-Reported Experience Measure (QTAC-PREM)	inadequate	doubtful	very good	0	0	Results show the expected correlation	+
		Joseph B, et al. J Trauma Acute Care Surg. 2017 Apr;82(4):722-727.	HACHPS scores correlate with clinical outcome parameters (complication scores, readmission, failure-to-rescue)	very good	N/A	very good	0	0	Results did not show the expected correlation	-
		Day MS, et al. J Healthc Qual. 2014 Nov-Dec;36(6):33-40.	Hospital-acquired conditions correlate with negative HCAHPS scores in an orthopedic hospital	0	0	0	adequate	very good	Results did not show the expected correlation	-
		Gupta A, et al. J Pain Res. 2009 Nov 13;2:157-64.	Adequate analgesic management results in higher HCAHPS scores	0	0	0	adequate	very good	Results show the expected correlation	+
		Kemp KA, et al. BMJ Open. 2016 Jul 1;6(7):e011242.	A patient-safety indicator correlates negatively with the HCAHPS global score and certain subscores	0	0	0	very good	adequate	Results show the expected negative correlation	+

		Sodhi N, et al. J Arthroplasty. 2019 Nov;34(11):2573-2579.	Patients readmitted within 30 or 90 days have lower HCAHPS scores than patients that have not been readmitted	0	0	0	very good	doubtful	Results show the expected negative correlation	+
5	QPPS	No studies on hypothesis testing for construct validity could be identified		0	0	0	0	0	0	?
6	DUQUE	No studies on hypothesis testing for construct validity could be identified		0	0	0	0	0	0	?
7	PEQ-G	No studies on hypothesis testing for construct validity could be identified		0	0	0	0	0	0	?
8	Schoenfelder et al.	Tonio Schoenfelder, et al. Int.J for Quality in Health Care, Volume 23, 5, 2011, 503–509	Higher age correlates inversely with the score	0	0	0	very good	very good	Results do not show the expected correlation	-
cancer-care specific PREMs										
1	EORTC IN-PATSAT32	Arraras JI, et al. Clin Transl Oncol. 2009 Apr;11(4):237-42.	Results of the IN-PATSAT32 correlate with results of the Oberst patients' perception of care quality and satisfaction scale	very good	adequate	doubtful	0	0	Results show the expected correlation	+
		Arraras JI, et al. Clin Transl Oncol. 2009 Apr;11(4):237-42.	Results of the IN-PATSAT32 correlate with results of the EORTC QLQ-INFO25	very good	very good	doubtful	0	0	Results show the expected correlation	+

1	EORTC IN-PATSAT32	Arraras JI, et al. Clin Transl Oncol. 2009 Apr;11(4):237-42.	Patients with higher Oberst "perception of care quality and satisfaction scale" scores have higher IN-PATSAT32 scores	0	0	0	very good	doubtful	Results show the expected correlation	+
		Arraras JI, et al. Clin Transl Oncol. 2009 Apr;11(4):237-42.	Patients rating "would recommend the hospital to others" higher, show higher global IN-PATSAT32 scores	0	0	0	very good	doubtful	Results show the expected correlation	+
		Zhang J, Xie S, Liu J, Sun W, Guo H, Hu Y, Gao X. Patient Prefer Adherence. 2014 Sep 18;8:1285-92.	Patients <58 years have higher scores than patients ≥58 years	0	0	0	very good	inadequate	Results show the expected correlation	?
		Zhang J, Xie S, Liu J, Sun W, Guo H, Hu Y, Gao X. Patient Prefer Adherence. 2014 Sep 18;8:1285-92.	Patients with lower educational standard give lower IN-PATSAT32 scores	0	0	0	adequate	inadequate	Results show the expected correlation	?
		Zhang L, Dai Z, Cheng S, et al. Support Care Cancer. 2015 Sep;23(9):2721-30.	Patients with lower educational standard give lower IN-PATSAT32 scores	0	0	0	adequate	inadequate	Results show the expected correlation	?
		Zhang L, Dai Z, Cheng S, et al. Support Care Cancer. 2015 Sep;23(9):2721-30.	Patients waiting >2 months for diagnostic procedures have lower scores than patients waiting <2monts	0	0	0	very good	inadequate	The expected correlation was not found	?
		Asadi-Lari M, et al. Support Care Cancer. 2015 Jul;23(7):1875-82.	Results of the IN-PATSAT32 correlate with results of the EORTC QLQ-INFO25	very good	very good	inadequate	0	0	Results show the expected correlation	+

		Aboshaiqah A, et al. Palliat Support Care. 2016 Dec;14(6):621-627.	Results of the IN-PATSAT32 correlate with results of the QOLQ-C15-PAL	very good	adequate	inadequate	0	0	Results correlate with physical function, emotional function and dem global health status	?
2	PSCC-G	Bokemeyer F, et al. BMC Health Serv Res. 2020 Oct 27;20(1):983	Higher scores in the REPERES subscale “patient satisfaction” correlate with higher PSCC-G scores	very good	very good	doubtful	0	0	Results show the expected correlation	+
		Bokemeyer F, et al. BMC Health Serv Res. 2020 Oct 27;20(1):983	Higher scores in the PASCOQ subscale “family and friends” correlate with higher PSCC-G scores	very good	very good	doubtful	0	0	Results show the expected correlation	+
		Bokemeyer F, et al. BMC Health Serv Res. 2020 Oct 27;20(1):983	Higher scores in the PASCOQ subscale “shared-decision making” correlate with higher PSCC-G scores	very good	very good	doubtful	0	0	Results show the expected correlation	+
		Bokemeyer F, et al. BMC Health Serv Res. 2020 Oct 27;20(1):983	Higher scores in the PASCOQ subscale “nursing and care” correlate with higher PSCC-G scores	very good	very good	doubtful	0	0	Results show the expected correlation	+
		Jean-Pierre P, Fiscella K, et al. Cancer. 2011;117(4):854–61	Higher results in the subscale “understanding and participation in care” of the CASE Cancer questionnaire correlate with higher PSCC scores	very good	very good	doubtful	0	0	Results show the expected correlation	+
		Jean-Pierre P, Fiscella K, et al. Cancer. 2011;117(4):854–61	Higher results in the subscale “seek and obtain information” of the CASE Cancer questionnaire correlate with higher PSCC scores	very good	very good	doubtful	0	0	Results show the expected correlation	+
2	PSCC-G	Jean-Pierre P, Fiscella K, et al. Cancer. 2011;117(4):854–61	Age, sex, and native language do not correlate with PSCC-G scores	0	0	0	very good	doubtful	Results show the expected lack of correlation	+

3	Danish National Cancer Patient Questionnaire	No studies on hypothesis testing for construct validity could be identified	0	0	0	0	0	0	0	?
4	SCCC	Esser P, et al. Health Qual Life Outcomes. 2021 May 17;19(1):147	Higher scores in the depression subscale correlate inversely with the global SCCC score	inadequate	inadequate	very good	0	0	Results show the expected negative correlation	+
		Esser P, et al. Health Qual Life Outcomes. 2021 May 17;19(1):147	Higher scores in the anxiety subscale correlate inversely with the global SCCC score	inadequate	inadequate	very good	0	0	Results show the expected negative correlation	+
		Esser P, et al. Health Qual Life Outcomes. 2021 May 17;19(1):147	Higher scores in the fatigue subscale correlate inversely with the global SCCC score	inadequate	inadequate	very good	0	0	Results show the expected negative correlation	+
		Esser P, et al. Health Qual Life Outcomes. 2021 May 17;19(1):147	Higher scores in the "total symptom load" subscale correlate inversely with the global SCCC score	inadequate	inadequate	very good	0	0	Results show the expected negative correlation	+

404 “+” = sufficient, ”-“ = insufficient, “?” = unclear

405 N/A: not applicable

406 0 : not reported

407

408

409

410 **Discussion**

411 In the current study numerous German language PREMs could be identified that were not
412 contained in previous publications [8,42]. This was due to publications in recent years as well
413 as due to the difficulty in identifying PREMs by database searches alone. Many German
414 PREMs were found by hand-searches. The current study uses for the first time the current
415 COSMIN guidelines for the assessment of PREMs [15]. Furthermore, by using a
416 comprehensive model of PC covering all dimensions of PC (S2 Table) a thorough analysis of
417 content validity of PREMs was possible for the first time. The study uncovers a deficit in
418 German language PREMs with adequate reliability and validity. The analysis of content
419 validity shows that there is no “perfect” PREM, but rather several available tools each with its
420 strengths and weaknesses. A context-specific application of German PREMs is mandatory and
421 several recommendations can be made.

422 **Recommendations**

423 Two PREMs cannot be recommended currently for use in German because of a lack of
424 validation of psychometric properties: the DUQUE questionnaire [27], as well as the German
425 translation of the *Danish National Cancer Patient Questionnaire* used by Rudolph et al. [21].
426 Depending on intended use, one of the remaining ten PREMs can be selected. Fig 2 shows a
427 schematic representation of the remaining PREMs within their intended area of use. The figure
428 can facilitate selection. In a next step, the results of this systematic review can be used to select
429 a PREM with sufficient psychometric properties (Table 5-10) and the necessary content (Table
430 3 and 4). For example, for cancer healthcare-specific PREMs the PSCC-G has significant better
431 psychometric properties than the SCC with its insufficient structural validity and lack of
432 assessment in many psychometric domains. Furthermore, when selecting a PREM the intended
433 area of application should to be considered [8]: is it intended as a reflection instrument for

434 patients or rather as a provider-specific evaluation instrument for internal use or as a
435 benchmarking instrument to compare different providers? In each case different content
436 dimensions (Table 3) and length of questionnaires (Table 2) are of interest.

437

438 **Fig 2. German language Patient-Reported Experience Measures (PREMs).** Overview of
439 German PREMs in their intended category of use

440

441 The following generic PREMs have been sufficiently evaluated in German: HCAHPS,
442 NORPEQ, PPE-15 and PEACS. For cancer care the EORTC IN-PATSAT32 and PSCC-G have
443 been adequately assessed and can currently be recommended. We were unable to identify a
444 surgery-specific PREM in the German language. However, even for the above mentioned
445 generic and cancer care-specific PREMs certain deficits need to be considered before use. The
446 HCAHPS for example, although showing sufficient psychometric properties in many areas,
447 exhibits deficits in its cross-cultural validation into German (Table 7) [34]. Some of its
448 demographic questions like “What is your race? Please select at least one.” were rated poorly
449 by native German speakers [34] and refer to its development in a different sociocultural
450 context. Therefore, an adaptation to the German-speaking sociocultural context seems
451 necessary. The PPE-15, one of the most frequently used PREMs worldwide, exhibits poor
452 structural validity, while covering many dimensions of PC (sufficient content validity). In
453 addition, for many psychometric properties of the PPE-15 no data could be found. We cannot
454 rule out that such data exist but has not been published by the Picker institutes or was not
455 identified by our search. The NORPEQ is an extensively studied PREM with adequate
456 psychometric properties. However, cross-cultural validation studies only exist for languages
457 other than German, although it has been used in a non-validated German translations [43]. One
458 of the most extensively evaluated generic PREMs is the PEACS questionnaire, that has been

459 developed in German with involvement of patients. It is a comprehensive questionnaire with
460 more than 50 questions covering many aspects of PC. Because of its length (Table 2) its
461 intended use is as a reflection instrument for patients and as an assessment tool for providers
462 rather than as benchmarking instrument. It is the only German generic PREM that covers not
463 only in-hospital aspects of patient experience, but also the transition into out-patient care.
464 Although the PEACS has sufficient psychometric properties in many areas, there is a lack of
465 data for test-retest, inter-rater and intra-rater reliability. The PSCC-G is a cancer care specific
466 PREM, that covers transition aspects of care. It has been built from a validated German
467 translation of an English questionnaire with additional questions from other languages (Table
468 2). The PSCC-G scored adequately in many psychometric domains, but data on test-retest,
469 inter-rater and intra-rater reliability are lacking.

470 **Limitations**

471 The study has several limitations. First, the search was limited to generic, surgery- and cancer
472 care specific PREMs, i.e., PREMs for other disciplines (e.g., internal medicine) as well as
473 PREMs for specific diseases were excluded. These PREMs can be found in the excluded
474 fulltext list (S7 Table). Another limitation could have been the search algorithm. The fact, that
475 many German-language PREMs were identified by hand-searching rather than the database
476 search, could be a hint that the search algorithm was not specific enough. However, the large
477 number of identified and screened articles indicates that our search was broad. Furthermore,
478 we were able to identify significantly more German-language PREMs than in previous reviews
479 [8,42]. Identifying PREMs in scientific databases is not easy. Contrary to PROMs no PREM-
480 specific taxonomy (e.g., MeSH term) exists for PREMs in common medical databases. As
481 pointed out, patient centredness and patient experience are only beginning to be clearly defined
482 (S1 Fig and S2 Table). The delineation to other concepts like patient satisfaction is not always
483 clear cut which makes building a search algorithm more difficult.

484 A main finding of the study is the lack of psychometric data for many of the included PREMs.
485 Frequently we were unable to find appropriate studies in accessible databases. However, many
486 PREMs have been developed and are implemented by independent or commercial institutions
487 or healthcare agencies. These institutions are often not scientifically driven and might not
488 publish all available psychometric data. Exception are the transparent development and
489 publication of data by the U.S. *Agency for Healthcare Research and Quality (AHRQ)*
490 (www.ahrq.gov/cahps/about-cahps/index.html) or the Swedish PREMs [18,44,45].

491 **Future research**

492 If PC is supposed to be more than a declaration of intent of healthcare politicians, it will require
493 the implementation of PREMs into everyday clinical practice via the following measures:

- 494 • As shown in our study, there is no comprehensive modular PREM system in the
495 German language comparable to other countries [23]. Many areas of healthcare (e.g.,
496 surgery) are not covered with available German-language PREMs. Consequently, the
497 development, translation and testing of new PREMs is necessary.
- 498 • The missing psychometric properties of currently available German-language PREMs
499 need to be evaluated.
- 500 • Most of the PREMs currently available are paper-based versions (Table 2). For broad
501 implementation and timely assessment in hospitals and doctor's offices electronic
502 PREMs (ePREM) seem necessary. For this purpose, paper-based PREMs will need to
503 be evaluated as digital versions and electronic systems will have to be developed and
504 implemented that adhere to local data safety regulations. An integration into available
505 hospital information systems is desirable, to facilitate the use in everyday clinical
506 practice.
- 507 • It is unclear which conclusions should be drawn from the results of PREM (sub)scales.
508 If providers adapt their service based on PREM results, there is little evidence-base to

509 guide such changes [46]. Individualized local measures may be implemented, but there
510 may also be standardized interventions, which can be tested in randomized-controlled
511 trials which might improve aspects of PC and subsequently PREM results. More
512 research is needed in this field.

513 There are two projects that should be mentioned in this context. First, the EORTC is currently
514 developing and testing the PATSAT-33, a cancer-care specific PREM that will not only cover
515 in-hospital patients, but also aspects of PC in out-patient settings as well as the transitional
516 period [47]. A phase IV validation study in several European countries is underway including
517 German-speaking countries. Second, the Hamburg-based ASPIRED project [48], is currently
518 developing a German-language PREM, that will cover all aspects of PC according to Scholl et
519 al. [5]. Both projects will close important evidence gaps.

520 **Conclusions**

521 This systematic review and qualitative analysis of the psychometric properties of German-
522 language PREMs, gives an overview of available generic and cancer-care specific PREMs.
523 Results can be used to choose valid and reliable PREMs according to the above-mentioned
524 recommendations. Furthermore, areas of future research are delineated.

525

526

527 **Author contributions**

528 AMi and GS designed the study. EK performed the database searches in coordination with
529 AMi. AMi and CDH screened titles and abstracts as well as full texts. AMi and CDH performed
530 the data extraction. AMi and CDH performed the qualitative analyses. AMi and GS drafted the
531 manuscript. All authors revised the manuscript critically. All authors read and approved the
532 manuscript.

533

534 **Acknowledgements**

535 The work is based on the Master thesis of AMi for the Master of Business Administration
536 (MBA) in Healthcare Management at the Wirtschaftsuniversität Wien (Vienna University of
537 Economics and Business) (2019-2021) for which GS acted as mentor and supervisor.

539 **References**

- 540 1. Bundesgesundheitsministerium. Ministerinterview mit Jens Spahn. 8 Jul 2020 [cited 20
541 Dec 2020]. Available:
542 <https://www.bundesgesundheitsministerium.de/gesundheitskompetenz/ministerinterview.html>
543
- 544 2. Bundesgesundheitsministerium. Patientenrechtegesetz. 2013 [cited 10 Sep 2020].
545 Available: <https://www.bundesgesundheitsministerium.de/service/begriffe-von-a-z/p/patientenrechtegesetz.html>
546
- 547 3. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st
548 Century. Washington (DC): National Academies Press (US); 2001. Available:
549 <https://pubmed.ncbi.nlm.nih.gov/25057539/>
- 550 4. Zeh S, Christalle E, Hahlweg P, Härter M, Scholl I. Assessing the relevance and
551 implementation of patient-centredness from the patients' perspective in Germany:
552 results of a Delphi study. *BMJ Open*. 2019;9: e031741. doi:10.1136/bmjopen-2019-
553 031741
- 554 5. Scholl I, Zill JM, Härter M, Dirmaier J. An integrative model of patient-centeredness - a
555 systematic review and concept analysis. *PloS One*. 2014;9: e107828.
556 doi:10.1371/journal.pone.0107828
- 557 6. BQS Institut. Picker Patient-centred care model. 2021 [cited 10 Sep 2020]. Available:
558 <https://www.bqs.de/picker-befragungen/51-patientenbefragung.php>
- 559 7. de Silva D. Measuring patient experience. Evidence scan No. 18. The Health
560 Foundation; 2013. Available: <https://www.health.org.uk/publications/measuring-patient-experience>
561
- 562 8. Bull C, Byrnes J, Hettiarachchi R, Downes M. A systematic review of the validity and
563 reliability of patient-reported experience measures. *Health Serv Res*. 2019;54: 1023–
564 1035. doi:10.1111/1475-6773.13187
- 565 9. Powers JH, Patrick DL, Walton MK, Marquis P, Cano S, Hobart J, et al. Clinician-
566 Reported Outcome Assessments of Treatment Benefit: Report of the ISPOR Clinical
567 Outcome Assessment Emerging Good Practices Task Force. *Value Health J Int Soc
568 Pharmacoeconomics Outcomes Res*. 2017;20: 2–14. doi:10.1016/j.jval.2016.11.005
- 569 10. Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al.
570 Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*.
571 2014;383: 166–175. doi:10.1016/S0140-6736(13)62227-8
- 572 11. Terwee, Prinsen, Chiarotto, deVet, Bouter, Alonso, et al. COSMIN methodology for
573 assessing the content validity of PROMs - A User Manual, version 1.0. COSMIN Study
574 Group; 2018. Available: <https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf>
575

- 576 12. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The
577 PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Int J*
578 *Surg Lond Engl*. 2021;88: 105906. doi:10.1016/j.ijisu.2021.105906
- 579 13. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The
580 COSMIN study reached international consensus on taxonomy, terminology, and
581 definitions of measurement properties for health-related patient-reported outcomes. *J*
582 *Clin Epidemiol*. 2010;63: 737–745. doi:10.1016/j.jclinepi.2010.02.006
- 583 14. Mokkink L, Prinsen C, Patrick D, Alonso J, Bouter L, de Vet H, et al. COSMIN
584 methodology for systematic reviews of Patient-Reported Outcome Measures. User
585 Manuel. Version 1.0. COSMIN Study Group; 2018.
- 586 15. Prinsen C.A.C., Mokkink L.B., Bouter L.M., Alonso J., Patrick D.L., de Vet H.C.W., et
587 al. COSMIN guideline for systematic reviews of patient-reported outcome measures.
588 *Qual Life Res*. 2018;27: 1147–1157. doi:10.1007/s11136-018-1798-3
- 589 16. Jenkinson C, Coulter A, Bruster S. The Picker Patient Experience Questionnaire:
590 development and validation using data from in-patient surveys in five countries. *Int J*
591 *Qual Health Care J Int Soc Qual Health Care*. 2002;14: 353–358.
592 doi:10.1093/intqhc/14.5.353
- 593 17. Brédart A, Razavi D, Delvaux N, Goodman V, Farvacques C, Van Heer C. A
594 comprehensive assessment of satisfaction with care for cancer patients. *Support Care*
595 *Cancer Off J Multinatl Assoc Support Care Cancer*. 1998;6: 518–523.
596 doi:10.1007/s005200050207
- 597 18. Wilde Larsson B, Larsson G. Development of a short form of the Quality from the
598 Patient’s Perspective (QPP) questionnaire. *J Clin Nurs*. 2002;11: 681–687.
599 doi:10.1046/j.1365-2702.2002.00640.x
- 600 19. Keller S, O’Malley AJ, Hays RD, Matthew RA, Zaslavsky AM, Hepner KA, et al.
601 Methods used to streamline the CAHPS Hospital Survey. *Health Serv Res*. 2005;40:
602 2057–2077. doi:10.1111/j.1475-6773.2005.00478.x
- 603 20. Oltedal S, Garratt A, Bjertnaes Ø, Bjørnsdottir M, Freil M, Sachs M. The NORPEQ
604 patient experiences questionnaire: data quality, internal consistency and validity
605 following a Norwegian inpatient survey. *Scand J Public Health*. 2007;35: 540–547.
606 doi:10.1080/14034940701291724
- 607 21. Rudolph C., Petersen G.S., Pritzkeleit R., Storm H., Katalinic A. The acceptance and
608 applicability of a patient-reported experience measurement tool in oncological care: a
609 descriptive feasibility study in northern Germany. *BMC Health Serv Res*. 2019;19: 786.
610 doi:10.1186/s12913-019-4646-4
- 611 22. Bokemeyer F., Lange-Drenth L., Jean-Pierre P., Schulz H., Bleich C. Psychometric
612 evaluation of the German version of the Patient Satisfaction with Cancer-related Care
613 questionnaire. *BMC Health Serv Res*. 2020;20: 983. doi:10.1186/s12913-020-05838-7
- 614 23. Esser P, Sautier L, Sarkar S, Schilling G, Bokemeyer C, Koch U, et al. Development
615 and preliminary psychometric investigation of the German Satisfaction with

- 616 Comprehensive Cancer Care (SCCC) Questionnaire. Health Qual Life Outcomes.
617 2021;19: 147. doi:10.1186/s12955-021-01784-y
- 618 24. Noest S, Ludt S, Klingenberg A, Glassen K, Heiss F, Ose D, et al. Involving patients in
619 detecting quality gaps in a fragmented healthcare system: development of a
620 questionnaire for Patients' Experiences Across Health Care Sectors (PEACS). *Int J Qual*
621 *Health Care J Int Soc Qual Health Care*. 2014;26: 240–249. doi:10.1093/intqhc/mzu044
- 622 25. Gehrlach C, Altenhöner T, Schwappach D. *Der Patients' Experience Questionnaire*.
623 Verlag Bertelsmann Stiftung; 2009.
- 624 26. Schoenfelder T, Klewer J, Kugler J. Determinants of patient satisfaction: a study among
625 39 hospitals in an in-patient setting in Germany. *Int J Qual Health Care*. 2011;23: 503–
626 509. doi:10.1093/intqhc/mzr038
- 627 27. DuQuE Project Consortium, Suñol R, Groene O. DUQuE Deepening our Understanding
628 of quality improvement in Europe. HOPE; 2011 [cited 4 Jul 2021]. Available:
629 <http://www.duque.eu/index.php?page=participating-countries>
- 630 28. Singer S, Götze H, Möbius C, Witzigmann H, Kortmann R-D, Lehmann A, et al.
631 Quality of care and emotional support from the inpatient cancer patient's perspective.
632 *Langenbecks Arch Surg*. 2009;394: 723–731. doi:10.1007/s00423-009-0489-5
- 633 29. Neijenhuijs KI, Jansen F, Aaronson NK, Bredart A, Groenvold M, Holzner B, et al. A
634 systematic review of the measurement properties of the European Organisation for
635 Research and Treatment of Cancer In-patient Satisfaction with Care Questionnaire, the
636 EORTC IN-PATSAT32. *Support Care Cancer Off J Multinatl Assoc Support Care*
637 *Cancer*. 2018;26: 2551–2560. doi:10.1007/s00520-018-4243-9
- 638 30. Jean-Pierre P, Fiscella K, Freund KM, Clark J, Darnell J, Holden A, et al. Structural and
639 reliability analysis of a patient satisfaction with cancer-related care measure: a multisite
640 patient navigation research program study. *Cancer*. 2011;117: 854–861.
641 doi:10.1002/cncr.25501
- 642 31. Defosse G, Mathoulin-Pelissier S, Ingrand I, Gasquet I, Sifer-Riviere L, Ingrand P, et
643 al. Satisfaction with care among patients with non-metastatic breast cancer:
644 development and first steps of validation of the REPERES-60 questionnaire. *BMC*
645 *Cancer*. 2007;7: 129. doi:10.1186/1471-2407-7-129
- 646 32. Kleeberg UR, Tews J-T, Ruprecht T, Höing M, Kuhlmann A, Runge C. Patient
647 satisfaction and quality of life in cancer outpatients: results of the PASQOC study.
648 *Support Care Cancer Off J Multinatl Assoc Support Care Cancer*. 2005;13: 303–310.
649 doi:10.1007/s00520-004-0727-x
- 650 33. Barrio-Cantalejo IM, Simón-Lorda P, Sánchez Rodríguez C, Molina-Ruiz A, Tamayo-
651 Velázquez MI, Suess A, et al. [Cross-cultural adaptation and validation of the Picker
652 Patient Experience Questionnaire-15 for use in the Spanish population]. *Rev Calid Asist*
653 *Organo Soc Espanola Calid Asist*. 2009;24: 192–206. doi:10.1016/j.cali.2009.02.003
- 654 34. Squires A, Bruyneel L, Aiken LH, Van den Heede K, Brzostek T, Busse R, et al. Cross-
655 cultural evaluation of the relevance of the HCAHPS survey in five European countries.

- 656 Int J Qual Health Care J Int Soc Qual Health Care. 2012;24: 470–475.
657 doi:10.1093/intqhc/mzs040
- 658 35. Singer S, Danker H, Meixensberger J, Briest S, Dietz A, Kortmann R-D, et al.
659 Structured multi-disciplinary psychosocial care for cancer patients and the perceived
660 quality of care from the patient perspective: a cluster-randomized trial. *J Cancer Res*
661 *Clin Oncol.* 2019;145: 2845–2854. doi:10.1007/s00432-019-03018-7
- 662 36. Brédart A, Bottomley A, Blazeby JM, Conroy T, Coens C, D’Haese S, et al. An
663 international prospective study of the EORTC cancer in-patient satisfaction with care
664 measure (EORTC IN-PATSAT32). *Eur J Cancer Oxf Engl* 1990. 2005;41: 2120–2131.
665 doi:10.1016/j.ejca.2005.04.041
- 666 37. Danish Cancer Society. Kræftpatienters behov og oplevelser med sundhedsvæsenet
667 under udredning og behandling Kræftens Bekæmpelses. ISBN: 978-87-7064-367-2;
668 2017. Available:
669 [https://www.cancer.dk/dyn/resources/File/file/1/6911/1513858648/2017_kraeftpatienters](https://www.cancer.dk/dyn/resources/File/file/1/6911/1513858648/2017_kraeftpatienters-s-behov-og-oplevelser-med-sundhedsvaesenet-under-udredning-og-behandling.pdf)
670 [s-behov-og-oplevelser-med-sundhedsvaesenet-under-udredning-og-behandling.pdf](https://www.cancer.dk/dyn/resources/File/file/1/6911/1513858648/2017_kraeftpatienters-s-behov-og-oplevelser-med-sundhedsvaesenet-under-udredning-og-behandling.pdf)
- 671 38. Pishkuhi MA, Salmaniyan S, Nedjat S, Zendedel K, Lari MA. Psychometric properties
672 of the Persian version of satisfaction with care EORTC-in-patsat32 questionnaire among
673 Iranian cancer patients. *Asian Pac J Cancer Prev APJCP.* 2014;15: 10121–10128.
674 doi:10.7314/apjcp.2014.15.23.10121
- 675 39. Obtel M, Serhier Z, Bendahhou K, Bennani M, Zidouh A, Benider A, et al. Validation
676 of EORTC IN-PATSAT 32 in Morocco: Methods and Processes. *Asian Pac J Cancer*
677 *Prev APJCP.* 2017;18: 1403–1409. doi:10.22034/APJCP.2017.18.5.1403
- 678 40. Kriston L, Scholl I, Hölzel L, Simon D, Loh A, Härter M. The 9-item Shared Decision
679 Making Questionnaire (SDM-Q-9). Development and psychometric properties in a
680 primary care sample. *Patient Educ Couns.* 2010;80: 94–99.
681 doi:10.1016/j.pec.2009.09.034
- 682 41. Coleman EA, Mahoney E, Parry C. Assessing the quality of preparation for posthospital
683 care from the patient’s perspective: the care transitions measure. *Med Care.* 2005;43:
684 246–255. doi:10.1097/00005650-200503000-00007
- 685 42. Beattie M, Murphy DJ, Atherton I, Lauder W. Instruments to measure patient
686 experience of healthcare quality in hospitals: a systematic review. *Syst Rev.* 2015;4: 97.
687 doi:10.1186/s13643-015-0089-0
- 688 43. Groene O, Arah OA, Klazinga NS, Wagner C, Bartels PD, Kristensen S, et al. Patient
689 Experience Shows Little Relationship with Hospital Quality Management Strategies.
690 *PLOS ONE.* 2015;10: e0131805. doi:10.1371/journal.pone.0131805
- 691 44. Wilde B, Larsson G, Larsson M, Starrin B. Quality of care. Development of a patient-
692 centred questionnaire based on a grounded theory model. *Scand J Caring Sci.* 1994;8:
693 39–48. doi:10.1111/j.1471-6712.1994.tb00223.x

- 694 45. Larsson G, Larsson BW, Munck IM. Refinement of the questionnaire “quality of care
695 from the patient’s perspective” using structural equation modelling. *Scand J Caring Sci.*
696 1998;12: 111–118.
- 697 46. Davidson KW, Shaffer J, Ye S, Falzon L, Emeruwa IO, Sundquist K, et al. Interventions
698 to improve hospital patient satisfaction with healthcare providers and systems: a
699 systematic review. *BMJ Qual Saf.* 2017;26: 596–606. doi:10.1136/bmjqs-2015-004758
- 700 47. Brédart A, Beaudeau A, Young T, Moura De Albuquerque Melo H, Arraras JI, Friend
701 L, et al. The European organization for research and treatment of cancer - satisfaction
702 with cancer care questionnaire: revision and extended application development.
703 *Psychooncology.* 2017;26: 400–404. doi:10.1002/pon.4127
- 704 48. Christalle E, Zeh S, Hahlweg P, Kriston L, Härter M, Scholl I. Assessment of patient
705 centredness through patient-reported experience measures (ASPIRED): protocol of a
706 mixed-methods study. *BMJ Open.* 2018;8: e025896. doi:10.1136/bmjopen-2018-
707 025896
- 708
- 709

710 **Supporting information**

711

712 **S1 Table.** Dimensions of patient centredness.

713

714 **S2 Table.** PRISMA Checklist

715

716 **S3 Text.** Search algorithm

717

718 **S4 Text.** Material and methods

719

720 **S5 Table.** Non-German PREMs

721

722 **S6 Table.** Content validity

Figure 1. PRISMA Flow Chart of included studies.

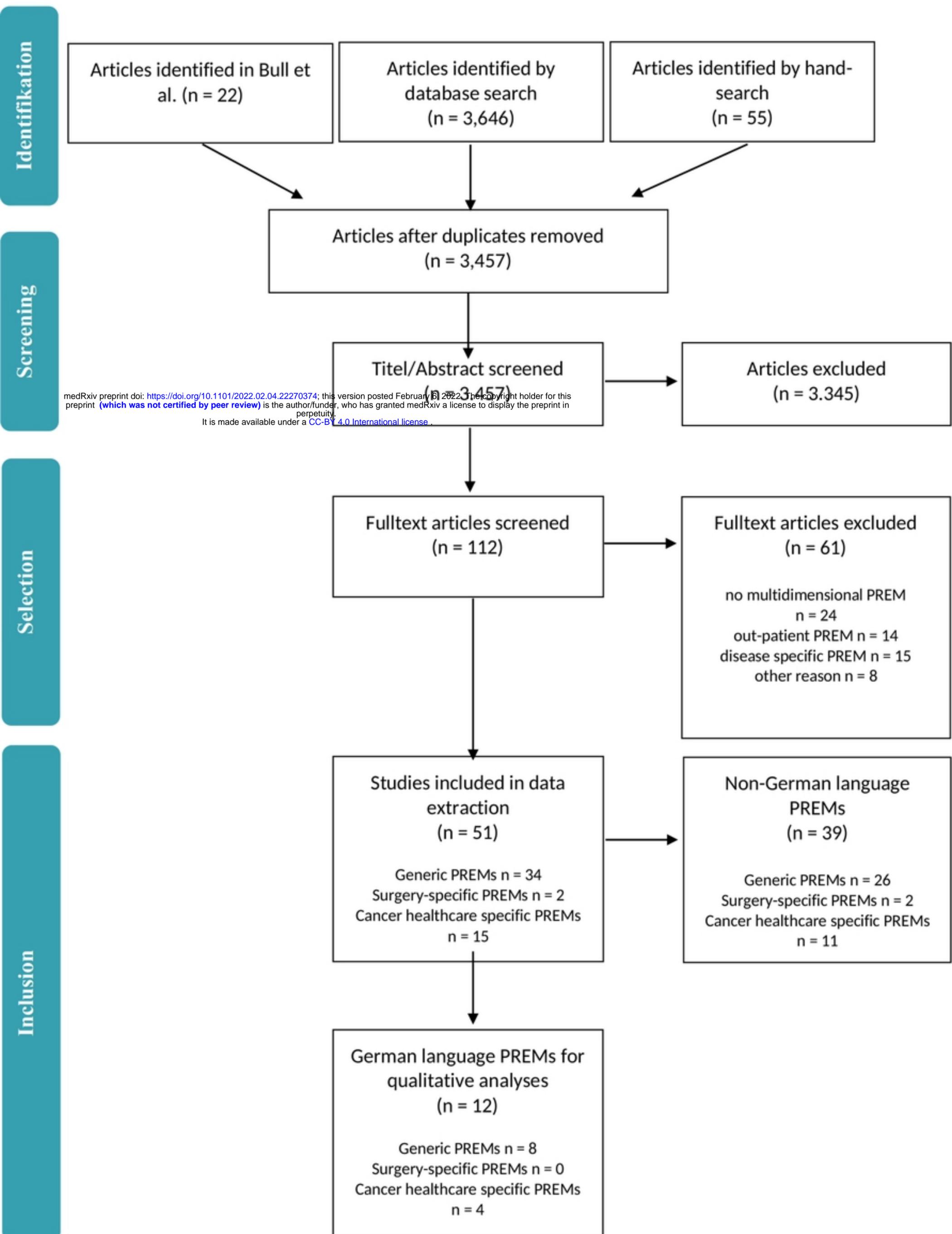


Figure 1

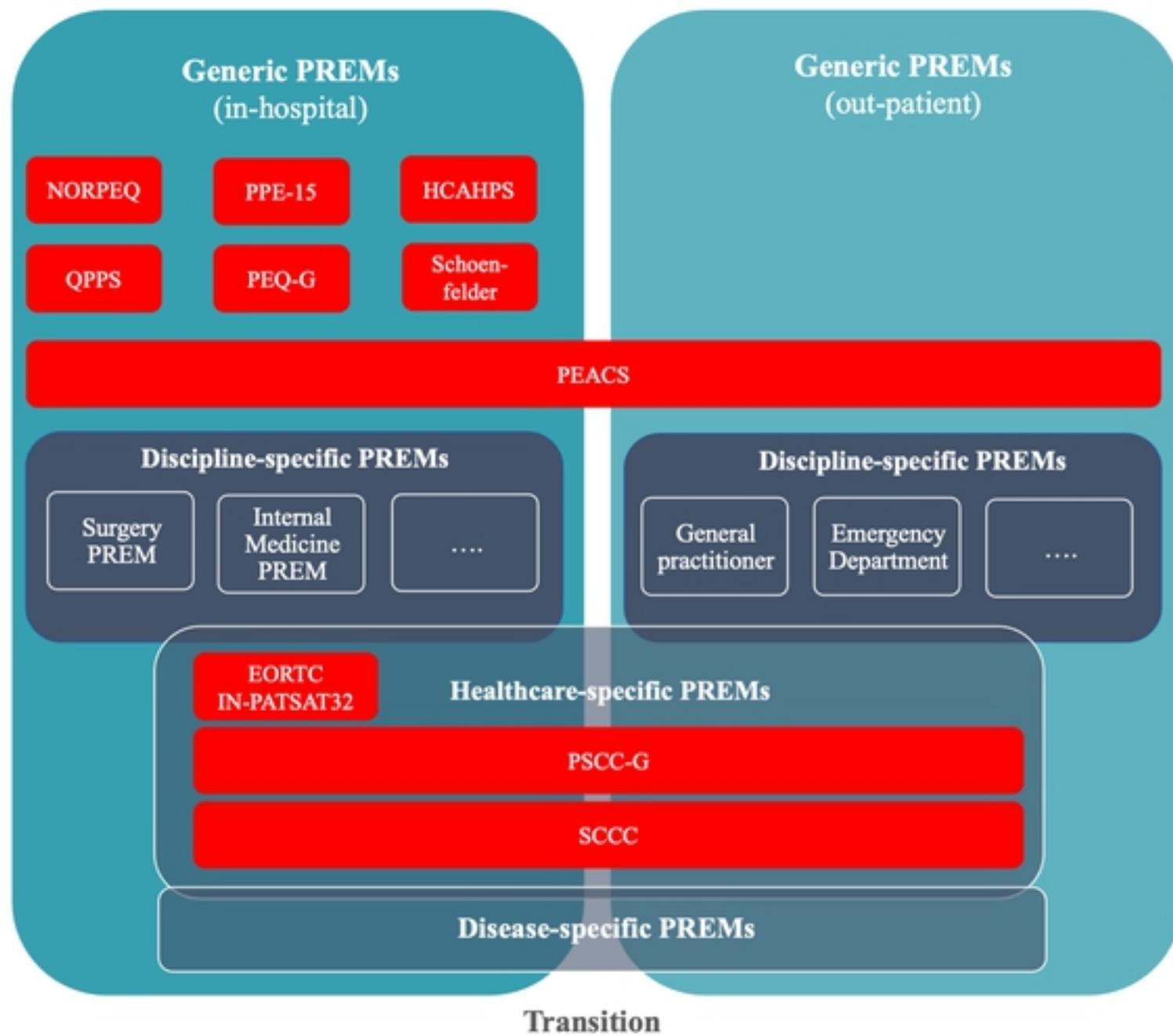


Figure 2