

ORIGINAL ARTICLE

# External validation and updating of prediction models for estimating the 1-year risk of low health-related quality of life in colorectal cancer survivors

Dóra Révész<sup>a</sup>, Sander M.J. van Kuijk<sup>b</sup>, Floortje Mols<sup>c,d</sup>, Fränzel J.B. van Duijnhoven<sup>e</sup>,  
Renate M. Winkels<sup>e</sup>, IJmert Kant<sup>f</sup>, Piet A. van den Brandt<sup>a,f</sup>, Luc J. Smits<sup>f</sup>,  
Stéphanie O. Breukink<sup>g</sup>, Ellen Kampman<sup>e</sup>, Sandra Beijer<sup>d</sup>, Matty P. Weijenberg<sup>a</sup>,  
Martijn J.L. Bours<sup>a,\*</sup>

<sup>a</sup>Department of Epidemiology, GROW – School for Oncology and Reproduction, Maastricht University, P. Debyeplein 1, 6200 MD Maastricht, The Netherlands

<sup>b</sup>Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre+, P. Debyelaan 25, PO Box 5800, Maastricht 6202 AZ, The Netherlands

<sup>c</sup>CoRPS – Center of Research on Psychology in Somatic Diseases, Department of Medical and Clinical Psychology, Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands

<sup>d</sup>Netherlands Comprehensive Cancer Organisation (IKNL), Godebaldkwartier 419, 3511 DT Utrecht, The Netherlands

<sup>e</sup>Division of Human Nutrition, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

<sup>f</sup>Department of Epidemiology, CAPHRI School for Public Health and Primary Care, Faculty of Health, Medicine and Life Sciences, Maastricht University, P. Debyeplein 1, 6200 MD Maastricht, The Netherlands

<sup>g</sup>Department of Surgery, Maastricht University Medical Centre, P. Debyelaan 25, 6229 HX Maastricht, The Netherlands

Accepted 30 September 2022; Published online 8 October 2022

## Abstract

**Objectives:** Timely identification of colorectal cancer (CRC) survivors at risk of experiencing low health-related quality of life (HRQoL) in the near future is important for enabling appropriately tailored preventive actions. We previously developed and internally validated risk prediction models to estimate the 1-year risk of low HRQoL in long-term CRC survivors. In this article, we aim to externally validate and update these models in a population of short-term CRC survivors.

**Study Design and Setting:** In a pooled cohort of 1,596 CRC survivors, seven HRQoL domains (global QoL, cognitive/emotional/physical/role/social functioning, and fatigue) were measured prospectively at approximately 5 months postdiagnosis (baseline for prediction) and approximately 1 year later by a validated patient-reported outcome measure (European Organization for Research and Treatment of Cancer Quality of Life Questionnaire—Core 30). For each HRQoL domain, 1-year scores were dichotomized into low vs. normal/high HRQoL. Performance of the previously developed multivariable logistic prediction models was evaluated (calibration and discrimination). Models were updated to create a more parsimonious predictor set for all HRQoL domains.

**Results:** Updated models showed good calibration and discrimination ( $AUC \geq 0.75$ ), containing a single set of 15 predictors, including nonmodifiable (age, sex, education, time since diagnosis, chemotherapy, radiotherapy, stoma, and comorbidities) and modifiable predictors (body mass index, physical activity, smoking, anxiety/depression, and baseline fatigue and HRQoL domain scores).

Conflict of interest: None declared.

**Funding:** This work was supported by a grant from the Alpe d'HuZes Foundation within the research program “Leven met kanker” of the Dutch Cancer Society (Grant number UM-2012-5653), and also partly by the Kankeronderzoekfonds Limburg as part of Health Foundation Limburg (Grant number 00005739). These funders only funded the design of the study, but played no role in the data collection, analysis, interpretation of results, or in the writing of the manuscript.

**Competing interests:** MJL Bours is a member of the editorial board of JCE, but was in no way involved in the handling, peer review process, and decision-making with regard to the manuscript.

**Ethics approval and consent to participate:** All procedures performed in studies involving human participants were approved by the Medical Ethics Committee of the University Hospital Maastricht and Maastricht University, The Netherlands, and the Committee on Research involving Human Subjects, region Arnhem-Nijmegen, The Netherlands. All patients signed informed consent.

**Availability of data and materials:** Datasets and code are available from the corresponding author on reasonable request.

\* Corresponding author. Department of Epidemiology, GROW – School for Oncology and Reproduction, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands, Tel.: +31 043 388 2903.

E-mail address: [m.bours@maastrichtuniversity.nl](mailto:m.bours@maastrichtuniversity.nl) (M.J.L. Bours).

**Conclusion:** Externally validated and updated prediction models performed well for estimating the 1-year risk of low HRQoL in CRC survivors within 6 months postdiagnosis. The impact of implementing the models in oncology practice to improve HRQoL outcomes in CRC survivors needs to be evaluated. © 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

**Keywords:** Colorectal cancer; Cancer survivors; Quality of life; Fatigue; Prediction models; External validation

## 1. Introduction

Globally the population of colorectal cancer (CRC) survivors is growing [1–3]. This growing population poses an increasing burden on health-care systems, as many survivors keep experiencing health problems and report decreased levels of health-related quality of life (HRQoL) in the years after diagnosis and treatment [4–6], with declines in levels of HRQoL mostly occurring in the first 6 months after diagnosis and treatment [7]. To provide appropriate and tailored care to CRC survivors in order to prevent declining HRQoL, identification of individual CRC survivors who have an increased risk of experiencing low HRQoL in the future is important.

Accurate and timely identification of high-risk individuals is challenging in oncology practice. Risk prediction models can aid health-care providers with identifying CRC survivors who are likely to experience low HRQoL in the future and who therefore are eligible for interventions aimed at promotion of their HRQoL. Prediction models have been developed to predict survival in CRC patients [8–10], which can provide input for the decision-making process regarding cancer treatment. When cancer treatments are completed, however, these models are not helpful anymore for predicting, for instance, the impact of late and long-term adverse treatment effects on HRQoL of CRC survivors. Prediction models specifically aimed at HRQoL estimation are needed for that purpose. Although previous studies have investigated associations of clinical, personal, lifestyle, and psychosocial factors with HRQoL in CRC survivors [11–13], these factors were not combined into prediction models to be used in oncology practice for predicting the risk of HRQoL declines after treatment.

Therefore, we have recently developed and internally validated prediction models to estimate the 1-year risk of low HRQoL in seven relevant domains (i.e., global quality of life; cognitive, emotional, physical, role, and social functioning; and fatigue) [14], using data from a large prospective cohort of long-term CRC survivors [15]. For model development, we used biopsychosocial HRQoL predictors that were selected based on evidence from previous association studies, summarized in an extensive systematic review [16]. Excellent predictive performance was demonstrated during model development and internal validation in CRC survivors on an average 5 years postdiagnosis [14], thereby covering the more long-term consequences of the disease and treatment on HRQoL. However, it is

necessary to externally validate the models in CRC survivors situated closer to the moment of diagnosis and treatment, since this is a more clinically relevant time frame for application of the models in oncology practice in order to timely identify individual CRC survivors at risk of experiencing low HRQoL earlier in the survivorship trajectory. If the predictive performance of the models is generalizable to short-term CRC survivors, this provides opportunities for tailoring of interventions, such as lifestyle and psychosocial interventions [17–24], aimed at safeguarding the HRQoL of high-risk individuals in the early post-treatment period. Moreover, the models can be used for selecting CRC survivors for studies evaluating the effectiveness of newly developed HRQoL interventions that could possibly be targeted at the modifiable predictors included in the models.

A common problem in the field of risk prediction modeling is that many models are being developed but only few models are externally validated, a crucial step to take before potential implementation into practice. Therefore, the purpose of the current study was twofold. First, we aimed to externally validate the previously developed prediction models for estimating the 1-year risk of low HRQoL in seven domains, using pooled data from two prospective cohorts of CRC survivors within the first year postdiagnosis. Second, we aimed to update the models to obtain a single set of HRQoL predictors to be used for all domains, producing more parsimonious models that are easier to implement in oncology practice, while still being useful for adequate risk prediction and identification of high-risk individual CRC survivors.

## 2. Methods

### 2.1. Study populations

For the current analyses, we pooled datasets of two prospective cohort studies (Figure 1): the “Energy for life after ColoRectal cancer” (EnCoRe) study and the “Colorectal cancer: Longitudinal, Observational study on Nutritional and lifestyle factors that may influence colorectal tumor recurrence, survival, and quality of life” (COLON) study.

#### 2.1.1. The EnCoRe study

We used data of  $N = 276$  CRC survivors from the EnCoRe study, which is described in more detail

**What is new?****Key findings**

- External validation of previously developed prediction models for estimating the 1-year risk of low health-related quality of life in seven domains showed adequate model performance (calibration and discrimination) in colorectal cancer survivors approximately 6 months postdiagnosis.
- Externally validated and updated models contained a total of 15 predictors, including sociodemographic and clinical predictors (age, sex, education, time since diagnosis, chemotherapy, radiotherapy, stoma, and comorbidities) and potentially modifiable lifestyle-related and psychosocial predictors (body mass index, physical activity, smoking, anxiety/depression, and baseline fatigue, and health-related quality of life domain scores).

**What this adds to what was known?**

- The externally validated and updated prediction models could assist clinicians to early identify individual colorectal cancer survivors who are at increased risk of experiencing low health-related quality of life in the near future, providing opportunities for timely preventive actions.

**What is the implication and what should change now?**

- As a next step toward possible implementation of the risk prediction models in oncology practice, the clinical impact of the models on improving or safeguarding health-related quality of life of colorectal cancer survivors needs to be evaluated.

elsewhere [25]. In short, EnCoRe is an ongoing multi-center prospective cohort study for which adult stages I–III CRC patients are being enrolled at diagnosis and followed up at 6 weeks, and at 6, 12, 24, and 60 months after treatment. Patients were recruited in three hospitals in the South-east of the Netherlands, and were excluded when diagnosed with stage IV CRC or in case of comorbidities obstructing successful participation (e.g., cognitive disorders such as Alzheimer disease). The study has been approved by the Medical Ethics Committee of the University Hospital Maastricht and Maastricht University, The Netherlands, and informed consent was obtained from all participants. Data from the 6-week (baseline) and 12-month post-treatment measurements (follow-up) collected between April 2012 and November 2016 were available for pooling (Figure 1).

**2.1.2. The COLON study**

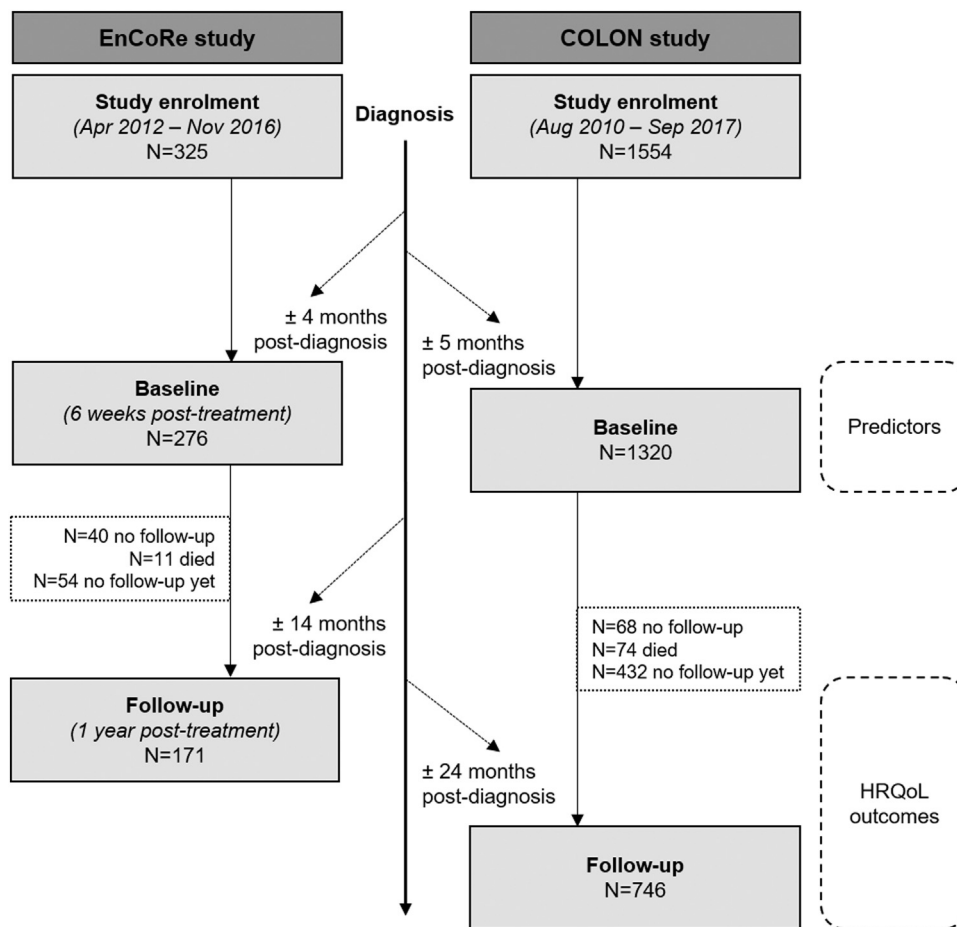
We also used data of  $N = 1,320$  CRC survivors from the COLON study, which is described in more detail elsewhere [26]. In short, COLON is an ongoing multi-center prospective cohort study among adult stages I–IV CRC patients recruited in eleven hospitals in the Netherlands. CRC patients were included at diagnosis and followed up at 6 months and at 2 years and 5 years after diagnosis. Patients were excluded when having a history of CRC or (partial) bowel resection, chronic inflammatory bowel disease, hereditary CRC syndromes, dementia or another mental condition obstructing participation. The study has been approved by the Committee on Research involving Human Subjects, region Arnhem-Nijmegen, The Netherlands, and informed consent was obtained from all participants. Data from the 6-month (baseline) and 2-year postdiagnosis measurements (follow-up) collected between August 2010 and September 2017 were available for pooling (Figure 1).

**2.2. Data collection****2.2.1. Health-related quality of life**

As reported previously [14], the prediction models aim to estimate at study baseline what the risk is of having low HRQoL in seven domains at follow-up approximately 1 year later. In both the EnCoRe and COLON study, HRQoL was measured at study baseline and follow-up with the European Organization for Research and Treatment of Cancer Quality of life Questionnaire—Core 30 (EORTC QLQ-C30, Version 3.0) [27]. Seven subscales were used to assess the following HRQoL domains: global QoL; cognitive, emotional, physical, role, and social functioning, and fatigue. For each subscale a sum score was calculated (0–100 points), with higher scores on global QoL and functioning scales representing better HRQoL, and higher scores on fatigue representing worse fatigue [27]. Since the previously developed prediction models estimate the risk of having low HRQoL at follow-up, scores of the separate HRQoL subscales were dichotomized into low vs. normal/high scores based on previously published medium-to-large minimal important deteriorations (MID) in the EORTC QLQ-C30 subscales [28], as also described in the development study [14]. For each domain, the low HRQoL group had a score  $\geq 1$  MID below the group mean score at baseline; the rest was included in the normal/high HRQoL group. Thus, individuals in the low HRQoL group either reported constantly low HRQoL scores at both baseline and follow-up, or experienced a clinically relevant deterioration from normal/high HRQoL scores at baseline to low scores at follow-up.

**2.2.2. Predictors**

Data collected within the EnCoRe and COLON cohorts on the predictors included in the previously developed



**Fig. 1.** Flowchart of participants from study enrollment to the baseline and follow-up measurements within the two prospective cohort studies used for the present analyses: the EnCoRe study and the COLON study. The baseline for prediction was approximately 4–5 months postdiagnosis in the pooled cohort. Health-related quality of life domains as outcomes for prediction were measured on average approximately 1 year after baseline. *Abbreviations:* EnCoRe, Energy for life after ColoRectal cancer; COLON, COlorectal cancer: Longitudinal, Observational study on Nutritional and lifestyle factors.

prediction models were used for the present analyses. Briefly, the original prediction models included the following baseline predictors, as described in more detail before [14]: age (years); sex; socio-economic status (high, medium, or low); number of comorbidities, as measured by the Self-Administered Comorbidity Questionnaire (none, 1, or 2+) [29]; time since diagnosis (years); stoma presence (yes/no); body mass index ( $\text{kg}/\text{m}^2$ ); adherence (yes/no) to Dutch physical activity guidelines of at least 150 minutes/week of moderate-to-vigorous physical activity measured by the Short QUEStionnaire to ASsess Health-enhancing physical activity [30]; anxiety and depressive symptoms measured by the Hospital Anxiety and Depression Scale [31]; baseline fatigue and HRQoL domain scores (EORTC QLQ-C30) [27]; chemotherapy (yes/no); radiotherapy (yes/no); tumor stage (TNM stages I–IV); current working status (yes/no); current smoking (yes/no); adherence (yes/no) to meat consumption recommendation according to the 2007 lifestyle guidelines of

the World Cancer Research Fund/American Institute for Cancer Research ( $<500$  gr meat per week [32]); social inhibition and negative affectivity measured with the Dutch 14-item Type D Personality Scale [33]; and subscale scores of micturition, pain, chemotherapy-related side effects, and stoma-related complaints, as well as a total score for gastrointestinal symptoms calculated by summing subscale scores of gastrointestinal problems, nausea/vomiting, constipation, defecation, and diarrhea (measured by the EORTC QLQ-CR29 module, version 2.1, with higher values indicating more complaints for all disease symptom scales) [34]. Education level was used as a proxy for socioeconomic status, which was included in the original prediction models, as educational level has been shown to have similar predictor properties in relation to health [35,36]. An overview of measurement instruments and methods for data collection in both the EnCoRe and COLON study is presented in [Supplementary Table 1](#), as well as a comparison with the instruments and methods used in the development study.

### 2.2.3. Missing data

Prior to data analyses, incomplete data on predictors and HRQoL outcomes were imputed with multiple imputation using the mice package in R (Version 1.0.136 – © 2009–2016 R Studio, Inc.) [37]. Some predictors from the original models were not measured in the EnCoRe and/or COLON cohorts (Supplementary Table 1) and therefore only partially or not available for the present analyses. Anxiety and depressive symptoms, working status, and disease symptom scales (micturition, chemotherapy-related side effects, stoma-related complaints, pain, and gastrointestinal complaints) were not available for the COLON cohort, and negative affectivity and social inhibition were not available in both EnCoRe and COLON. To deal with these missing data, we added data from the cohort used for developing the models [14] to proceed with the multiple imputation, as recommended in simulation studies to minimize bias [38,39]. In the multiple imputation procedure, we also took into account differences in time since diagnosis between the development cohorts (PROFILES) and the external validation cohorts (EnCoRe and COLON) by including time interactions.

### 2.3. Statistical analyses

To estimate the absolute risk of scoring low on seven HRQoL domains, multivariable logistic regression analyses were performed with the rms package in R [40]. The analyses consisted of two parts: (1) external validation of the originally developed and internally validated models, and (2) model updating. In both parts of the analyses, models were recalibrated when needed.

In the first part of the analyses, the prediction models with their original regression parameters were externally validated by evaluating their predictive performance in the pooled EnCoRe and COLON dataset (model 1A). The models were then recalibrated by updating the intercepts only (model 1B), to make the average predicted probability equal to the observed overall event rate (i.e., the prevalence of low HRQoL at follow-up in the pooled EnCoRe and COLON cohorts), or by updating both the intercept and the slope (model 1C). For the latter, a logistic regression model was fitted with the linear predictor, which is the sum of all regression coefficients multiplied by their predictor variable value, as the only covariate to estimate the new intercept and a shrinkage factor for adjusting all regression coefficients.

In the second part of the analyses, the models (1A–1C) for each of the separate HRQoL domains showing the best calibration were selected for further updating to obtain a single set of predictors for the seven HRQoL domains. The goal was to downsize the prediction models to create more parsimonious models that included less predictors and would be easier to implement in practice, without loss of predictive power. Therefore, we examined to which extent removal of predictors altered model performance, while in any case keeping the predictors for which the

strongest evidence was reported [16] in the updated models (model 2A). Lastly, the downsized models were recalibrated again by updating the intercept only (model 2B) or both the intercept and slope (model 2C), as described above.

In both parts of the analyses, evaluation of model performance included assessment of discrimination, calibration, overall performance, and classification. Discrimination reflects ability of a model to distinguish between individuals with low vs. normal/high HRQoL based on estimated risks, as quantified by the area under the Receiver Operator Characteristic curve (AUC, with  $AUC > 0.8$  indicating good discrimination) [41]. Calibration is a measure of the agreement between predicted probabilities and observed relative frequencies of low HRQoL, and is assessed by calibration-in-the-large (the average difference between predicted probabilities and observed relative frequencies, which should be small for adequately calibrated models), the Hosmer–Lemeshow goodness-of-fit test (H–L, with  $P > 0.05$  indicating adequate calibration), and by visually inspecting calibration plots showing agreement between predicted risk and observed prevalence of low HRQoL [42]. We assessed overall model performance with Nagelkerke's  $R^2$ , a measure of predictive strength (range 0–1) with higher values for better performance, and Brier scores that determine measures of model accuracy (0–0.25), with lower scores for better accuracy [43]. In addition, for a range of predicted probabilities, sensitivity and specificity were determined as measures of classification for the updated models after recalibration. Sensitivity reflects the percentage of true-positive predictions given low HRQoL, and specificity reflects the percentage of true-negative predictions given no low HRQoL. Optimal threshold probabilities to be used as cutoffs for classification were defined for each model based on a high sensitivity ( $\geq 80\%$ ), as also described in the development study [14]. High sensitivity was prioritized over high specificity, since we deemed false-negative misclassifications more problematic than false-positive misclassifications when considering the nonhazardous and noninvasive nature of interventions aimed at HRQoL promotion.

All analyses were performed using R software (R Foundation for Statistical Computing Platform, Version 1.0.136 © 2009–2016 RStudio, Inc.). The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis statement was used for the analyses and reporting [44,45].

## 3. Results

### 3.1. Population characteristics

CRC survivors participating in the EnCoRe and COLON studies were on average 66 years old and 5.2 months post-diagnosis, and 35% was female (Table 1). Whereas the two

**Table 1.** Predictors measured at baseline in the two nonimputed sets of the EnCoRe ( $N = 276$ ) and COLON study ( $N = 1,320$ ) separately and in the pooled dataset ( $N = 1,596$ )

	EnCoRe study <sup>a</sup>		COLON study <sup>a</sup>		EnCoRe + COLON <sup>a</sup>	
	$(N = 276)$		$(N = 1,320)$		$(N = 1,596)$	
	Total $N$	$N$ (%) <sup>b</sup>	Total $N$	$N$ (%) <sup>b</sup>	Total $N$	$N$ (%) <sup>b</sup>
<b>Sociodemographic factors</b>						
Age (mean yr, SD)	276	66.8 (9.4)	1,320	66.2 (8.7)	1,596	66.3 (8.9)
Sex (female)	276	88 (31.9)	1,320	475 (36.0)	1,596	563 (35.3)
Work status (not working)	267	192 (71.9)	-	-	267	192 (71.9)
Education level	275		1,319		1,594	
High		95 (34.5)		504 (38.2)		599 (37.6)
Medium		110 (40.0)		490 (37.1)		600 (37.6)
Low		70 (25.5)		325 (24.6)		395 (24.8)
<b>Clinical Factors</b>						
Number of comorbidities	276		1,294		1,570	
None		59 (21.4)		898 (69.4)		957 (61.0)
1		70 (25.4)		304 (23.5)		374 (23.8)
$\geq 2$		147 (53.3)		92 (7.1)		239 (15.2)
Time since diagnosis (mean yr, SD)	274	4.4 (3.1)	1,320	5.3 (1.0)	1,594	5.2 (1.6)
Tumor stage	265		1,060		1,325	
I		79 (29.8)		269 (25.4)		348 (26.3)
II		65 (24.5)		313 (29.5)		378 (28.5)
III		121 (45.7)		400 (37.7)		521 (39.3)
IV		-		78 (7.4)		78 (5.9)
Stoma present	266	87 (32.7)	1,320	322 (24.4)	1,586	409 (25.8)
Chemotherapy	276	109 (39.5)	1,319	462 (35.0)	1,595	571 (35.8)
Radiotherapy	276	78 (28.3)	1,319	226 (17.1)	1,595	304 (19.0)
Stoma-related complaints (mean, SD) <sup>c</sup>	265	8.3 (15.8)	-	-	1,263	1.8 (8.0)
Pain (mean, SD) <sup>c</sup>	267	18.4 (23.2)	1,268	13.4 (21.8)	1,535	14.3 (22.1)
Micturition (mean, SD) <sup>c</sup>	265	26.1 (19.2)	-	-	265	26.1 (19.2)
Chemotherapy-related side effects (mean, SD) <sup>c</sup>	266	14.2 (16.4)	-	-	266	14.2 (16.4)
Gastro-intestinal complaints sum score (mean, SD) <sup>c</sup>	175	52.6 (50.1)	-	-	175	52.6 (50.1)
<b>Body composition and lifestyle factors</b>						
Body mass index (mean in kg/m <sup>2</sup> , SD)	273	27.7 (4.5)	1,316	26.4 (3.8)	1,589	26.6 (4.0)
Smoking status (smokers)	266	25 (9.4)	1,320	94 (7.1)	1,586	119 (7.5)
Physical activity (adherence) <sup>d</sup>	268	162 (60.4)	1,293	846 (65.4)	1,561	1,008 (64.6)
Meat consumption (nonadherence) <sup>e</sup>	211	140 (66.4)	1,203	666 (55.4)	1,414	806 (57.0)
<b>Psychological factors</b>						
Anxiety score (mean, SD)	266	3.5 (3.3)	-	-	266	3.5 (3.3)
Depression score (mean, SD)	266	3.8 (3.6)	-	-	266	3.8 (3.6)

**Abbreviations:** EnCoRe, Energy for life after ColoRectal cancer; COLON, COlorectal cancer: Longitudinal, Observational study on Nutritional and lifestyle factors; SD, standard deviation.

<sup>a</sup> Baseline was 6 wk after the end of treatment in EnCoRe, whereas it was ~6 mo after diagnosis in COLON.

<sup>b</sup> Percentages are shown of the total number of participants with valid data.

<sup>c</sup> For all complaints a higher score (range 0–100) represents more complaints. A sum score (range 0–500) is created by summing up separate scores of gastrointestinal problems, nausea/vomiting, constipation, defecation, and diarrhea.

<sup>d</sup> Adherence to physical activity guidelines defined as  $\geq 30$  min/day of moderate-to-vigorous intensity physical activity on  $\geq 5$  days/wk.

<sup>e</sup> Adherence to WCRF/AICR meat consumption guidelines, defined as eating  $\leq 500$  g meat per week.

studies were similar regarding most predictors, participants of the EnCoRe study had a slightly higher number of comorbidities, more frequently a stoma, more often received

radiotherapy, and were less often adhering to physical activity guidelines. In the pooled cohort ( $N = 1,596$ ), 11–19% of participants were categorized as having low

**Table 2.** Health-related quality of life (HRQoL) domains at baseline and follow-up of the two nonimputed sets of the EnCoRe ( $N = 276$ ) and COLON study ( $N = 1,320$ ) separately and in the pooled dataset ( $N = 1,596$ )

	EnCoRe study		COLON study		EnCoRe + COLON	
	<i>N</i>	Mean (SD) / <i>N</i> (%) <sup>a</sup>	<i>N</i>	Mean (SD) / <i>N</i> (%) <sup>a</sup>	<i>N</i>	Mean (SD) / <i>N</i> (%) <sup>a</sup>
<b>Global QoL<sup>b</sup></b>						
Baseline (mean, SD)	267	74.1 (18.6)	1,284	74.7 (19.1)	1,551	74.6 (19.0)
Follow-up (mean, SD)	168	77.8 (17.7)	746	80.3 (17.0)	914	79.8 (17.1)
Low HRQoL: $\geq 1$ MID (=10) below baseline mean ( <i>N</i> , %) <sup>c</sup>		24 (14.3)		95 (12.7)		119 (13.0)
Consistently low ( <i>N</i> , %) <sup>d</sup>		12 (50.0)		48 (50.5)		60 (50.4)
Deteriorating ( <i>N</i> , %) <sup>d</sup>		12 (50.0)		47 (49.5)		59 (49.6)
<b>Cognitive functioning<sup>b</sup></b>						
Baseline (mean, SD)	267	86.2 (21.1)	1,275	85.9 (18.4)	1,542	86.0 (18.9)
Follow-up (mean, SD)	168	87.5 (19.4)	744	88.0 (17.3)	912	87.9 (17.7)
Low HRQoL: $\geq 1$ MID (=7) below baseline mean ( <i>N</i> , %) <sup>c</sup>		30 (17.9)		137 (18.4)		167 (18.3)
Consistently low ( <i>N</i> , %) <sup>d</sup>		22 (73.3)		77 (56.2)		99 (59.3)
Deteriorating ( <i>N</i> , %) <sup>d</sup>		8 (26.7)		52 (38.0)		60 (35.9)
<b>Emotional functioning<sup>b</sup></b>						
Baseline (mean, SD)	267	87.5 (18.5)	1,276	85.8 (17.6)	1,543	86.1 (17.8)
Follow-up (mean, SD)	168	88.1 (18.5)	741	87.5 (17.1)	909	87.6 (17.4)
Low HRQoL: $\geq 1$ MID (=12) below baseline mean ( <i>N</i> , %) <sup>c</sup>		36 (21.4)		125 (16.9)		161 (17.7)
Consistently low ( <i>N</i> , %) <sup>d</sup>		25 (69.4)		61 (48.8)		86 (53.4)
Deteriorating ( <i>N</i> , %) <sup>d</sup>		11 (30.6)		55 (44.0)		66 (41.0)
<b>Physical functioning<sup>b</sup></b>						
Baseline (mean, SD)	267	76.5 (19.6)	1,262	82.0 (17.9)	1,529	81.0 (18.4)
Follow-up (mean, SD)	168	82.8 (19.5)	730	86.6 (15.2)	898	85.9 (16.2)
Low HRQoL: $\geq 1$ MID (=10) below baseline mean ( <i>N</i> , %) <sup>c</sup>		32 (19.0)		110 (15.1)		142 (15.8)
Consistently low ( <i>N</i> , %) <sup>d</sup>		25 (78.1)		59 (53.6)		84 (59.2)
Deteriorating ( <i>N</i> , %) <sup>d</sup>		7 (21.9)		46 (41.8)		53 (37.3)
<b>Role functioning<sup>b</sup></b>						
Baseline (mean, SD)	267	70.5 (28.1)	1,276	74.6 (29.4)	1,543	73.9 (29.2)
Follow-up (mean, SD)	168	84.1 (23.6)	740	86.3 (23.0)	908	85.9 (23.1)
Low HRQoL: $\geq 1$ MID (=14) below baseline mean ( <i>N</i> , %) <sup>c</sup>		19 (11.3)		80 (10.8)		99 (10.9)
Consistently low ( <i>N</i> , %) <sup>d</sup>		12 (63.2)		41 (51.3)		53 (53.5)
Deteriorating ( <i>N</i> , %) <sup>d</sup>		7 (36.8)		38 (47.5)		45 (45.5)
<b>Social functioning<sup>b</sup></b>						
Baseline (mean, SD)	267	81.5 (22.2)	1,282	82.5 (22.5)	1,549	82.3 (22.5)
Follow-up (mean, SD)	168	90.6 (18.6)	746	89.3 (18.4)	914	89.5 (18.4)
Low HRQoL: $\geq 1$ MID (=11) below baseline mean ( <i>N</i> , %) <sup>c</sup>		27 (16.1)		145 (19.4)		172 (18.8)
Consistently low ( <i>N</i> , %) <sup>d</sup>		23 (85.2)		74 (51.0)		97 (56.4)
Deteriorating ( <i>N</i> , %) <sup>d</sup>		4 (14.8)		67 (46.2)		71 (41.3)
<b>Fatigue<sup>b</sup></b>						
Baseline (mean, SD)	267	28.9 (23.1)	1,258	28.0 (25.1)	1,525	28.2 (24.8)
Follow-up (mean, SD)	168	21.6 (22.9)	729	18.3 (20.1)	897	18.9 (20.7)
Low HRQoL: $\geq 1$ MID (=10) above baseline mean ( <i>N</i> , %) <sup>c</sup>		32 (19.0)		100 (13.7)		132 (14.7)

(Continued)

Table 2. Continued

	EnCoRe study		COLON study		EnCoRe + COLON	
	N	Mean (SD) / N (%) <sup>a</sup>	N	Mean (SD) / N (%) <sup>a</sup>	N	Mean (SD) / N (%) <sup>a</sup>
Consistently low (N, %) <sup>d</sup>		22 (68.8)		52 (52.0)		74 (56.1)
Deteriorating (N, %) <sup>d</sup>		10 (31.3)		43 (43.0)		53 (40.2)

*Abbreviations:* EnCoRe, Energy for life after ColoRectal cancer; HRQoL, health-related quality of life; COLON, COlorectal cancer: Longitudinal, Observational study on Nutritional and lifestyle factors; SD, standard deviation.

<sup>a</sup> Percentages are shown of the total number of participants with valid data.

<sup>b</sup> Higher scores (range 0–100) on global QoL and functioning domains represent better HRQoL, whereas higher fatigue scores represent worse fatigue complaints.

<sup>c</sup> Persons were classified as having “low HRQoL” when their follow-up score differed  $\geq 1$  minimal important deterioration (MID) [19] from baseline mean.

<sup>d</sup> Persons had consistently low HRQoL when both their baseline and follow-up scores were  $\geq 1$  MID above/below the baseline mean. Persons had deteriorating HRQoL when they decreased from normal/high HRQoL at baseline to a follow-up score  $\geq 1$  MID below/above the baseline group mean (below mean for global quality of life and functioning domains; above mean for fatigue). The number of subjects with consistently low and deteriorating HRQoL do not always add up to the total number in the low HRQoL group due to missings at baseline.

HRQoL at follow-up in the seven domains (Table 2), with slightly more participants showing consistently low HRQoL (50–59%) than deteriorating HRQoL between baseline and follow-up (36–50%). The EnCoRe and COLON study were similar with regard to the prevalence of low HRQoL at follow-up (Table 2).

Of all participants, 868 (54.4%) had complete HRQoL data at follow-up. Participants who had missing HRQoL data at follow-up ( $N = 728$ ) were more often diagnosed with stage IV disease (10% vs. 3%) and less often with stage I disease (24% vs. 28%,  $P = 0.004$ ), and less frequently had received radiotherapy (17% vs. 21%,  $P = 0.03$ ). There were no differences regarding baseline HRQoL scores between the participants with missing vs. complete data. Differences between the development cohort and the external validation cohorts are shown in Supplementary Tables 2 and 3

### 3.2. External validation

First, the original models (model 1A) were externally validated, yielding good discrimination (AUC: 0.77–0.84), sufficient measures of overall performance, and reasonable calibration (Table 3). The calibration plots showed that most models overestimated the risk of CRC survivors to score low on HRQoL domains (Supplementary Figure 1). After recalibration (i.e., updating the intercepts and slopes of the models; models 1B and 1C, respectively), the calibration of the separate models improved, as the predictions were in closer agreement with observed relative frequencies of low HRQoL shown in the calibration plots (Supplementary Figure 1). Models with updated intercepts and slopes (model 1C) showed the best calibration for most HRQoL domains, which were therefore selected for the model updating.

### 3.3. Model updating

In order to obtain a single set of predictors for the seven HRQoL domains (model 2A), we removed the following predictors from model 1C: all EORTC QLQ-CR29

symptom scales, negative affectivity and social inhibition scores, tumor stage, adherence to meat consumption guidelines, and current working status. The updated models thus included the following unified set of 15 predictors: age, sex, education, comorbidities, time since diagnosis, stoma, body mass index, physical activity, anxiety, depression, chemotherapy, radiotherapy, smoking, and baseline fatigue and HRQoL scores (Table 4). Discrimination did not change considerably following removal of predictors (AUC: 0.77–0.85), and overall model performance measures were similar to those of the original models (Table 3). The calibration plots showed that the models now underestimated the risk of scoring low on HRQoL (Supplementary Figure 1). Updating the intercept (model 2B) improved the calibration plots for the separate HRQoL domains, with most predictions now lying close to the diagonal indicating good agreement between predicted probabilities and relative frequencies of low HRQoL. As updating the slope in addition to the intercept (model 2C) did not further improve the calibration, model 2B was chosen as the final updated model.

### 3.4. Measures of classification

Sensitivities and specificities of the updated models (model 2B) were derived at different thresholds of predicted risks of low HRQoL (plotted in Supplementary Figure 2). Sensitivity  $\geq 80\%$  was reached with the following optimal threshold probabilities as cut-off for positive predictions for the separate HRQoL domains (indicated by the gray areas in Supplementary Figure 2): 5% for role functioning; 5–10% for global QoL, cognitive functioning, emotional functioning, and fatigue; 10–15% for social functioning; and 5–15% for physical functioning.

## 4. Discussion

We externally validated previously developed prediction models for estimating the 1-year risk of low HRQoL in



**Table 3.** Model performance measures of prediction models for seven domains of health-related quality of life: performance measures of original models and updated models after external validation and recalibration are presented

	Global quality of life	Cognitive functioning	Emotional functioning	Physical functioning	Role functioning	Social functioning	Fatigue
Model 1A: Original models							
AUC [95% CI] <sup>a</sup>	0.78 [0.75–0.81]	0.83 [0.81–0.86]	0.82 [0.80–0.85]	0.85 [0.82–0.87]	0.77 [0.73–0.80]	0.77 [0.74–0.80]	0.81 [0.78–0.83]
Nagelkerke's R <sup>2b</sup>	0.200	0.344	0.312	0.352	0.174	0.213	0.253
Brier score <sup>c</sup>	0.098	0.112	0.107	0.108	0.089	0.135	0.107
Calibration-in-the-large <sup>d</sup>	0.181	–0.002	–0.003	–0.005	0.196	0.142	0.097
H–L test P-value <sup>e</sup>	<0.001	0.42	0.18	<0.001	<0.001	0.33	<0.001
Model 1B: Original models after recalibration: updating of intercept							
Calibration-in-the-large <sup>d</sup>	0.000	0.000	0.000	0.000	0.000	0.000	0.000
H–L test P-value <sup>e</sup>	<0.001	0.42	0.21	<0.001	0.003	0.003	<0.001
Model 1C: Original models after recalibration: updating of intercept and slope							
Calibration-in-the-large <sup>d</sup>	–0.013	0.044	–0.017	0.061	–0.043	–0.041	–0.103
H–L test P-value <sup>e</sup>	0.33	0.10	0.86	<0.001	0.002	0.92	<0.001
Model 2A: Updated models							
AUC [95% CI] <sup>a</sup>	0.78 [0.75–0.81]	0.83 [0.80–0.85]	0.82 [0.79–0.85]	0.85 [0.83–0.87]	0.76 [0.73–0.80]	0.75 [0.72–0.78]	0.80 [0.77–0.82]
Nagelkerke's R <sup>2b</sup>	0.196	0.332	0.299	0.357	0.170	0.192	0.233
Brier score <sup>c</sup>	0.099	0.114	0.108	0.107	0.089	0.137	0.109
Calibration-in-the-large <sup>d</sup>	–0.039	–0.011	–0.069	0.047	–0.056	–0.120	–0.120
H–L test P-value <sup>e</sup>	0.001	0.63	<0.001	0.001	<0.001	<0.001	<0.001
Model 2B: Updated models after recalibration: updating of intercept							
Calibration-in-the-large <sup>d</sup>	0.000	0.000	0.000	0.000	0.000	0.000	0.000
H–L test P-value <sup>e</sup>	0.34	0.68	0.52	0.28	0.86	0.28	0.13
Model 2C: Updated models after recalibration: updating of intercept and slope							
Calibration-in-the-large <sup>d</sup>	0.032	–0.018	0.017	–0.004	0.037	0.098	0.082
H–L test P-value <sup>e</sup>	0.10	0.57	0.41	0.35	0.05	<0.001	<0.001

<sup>a</sup> AUC = Area under the Receiver Operator Characteristic curve; 95% confidence intervals (AUC  $\geq$  0.80 indicates good discrimination).

<sup>b</sup> Nagelkerke's R<sup>2</sup> is a measure of explained variance, ranging from 0 to 1 (higher is better).

<sup>c</sup> Brier score is a measure of model accuracy, ranging from 0 (perfect) to 0.25 (worthless accuracy).

<sup>d</sup> Calibration-in-the-large is the difference between the predicted probability and the observed relative frequency of low HRQoL.

<sup>e</sup> H–L = Hosmer–Lemeshow goodness-of-fit test is an indicator of calibration (agreement between observed and predicted values);  $P > 0.05$  represents well-calibrated model (i.e., nonsignificant disagreement between observed and predicted values).

seven domains in CRC survivors within 6 months after diagnosis, using pooled data from two prospective cohorts. In addition, the models were updated by removing several predictors to yield a single predictor set for all HRQoL domains, producing downsized models that are more parsimonious and therefore easier to implement in oncology practice. All updated models had satisfactory predictive power after recalibration for identifying individual CRC survivors at increased risk of low future HRQoL in the seven domains, as indicated by adequate measures of overall model performance, discrimination, calibration, and classification. The updated models included 15 predictors

in total, containing both nonmodifiable (age, sex, education, time since diagnosis, chemotherapy, radiotherapy, stoma, and comorbidities) and modifiable predictors (body mass index, physical activity, smoking, anxiety and depression, and baseline fatigue and HRQoL domain scores).

The prediction models that had previously been developed in long-term CRC survivors (approximately 5 years postdiagnosis) [14] were now externally validated and updated in short-term CRC survivors (approximately 5 months postdiagnosis), to assess whether models were generalizable to a time frame situated closer to the diagnosis and treatment during which larger declines in HRQoL are

**Table 4.** Regression coefficients and odds ratios of included predictors in models 2B for seven domains of health-related quality of life

	Global quality of life		Cognitive functioning		Emotional functioning		Physical functioning		Role functioning		Social functioning		Fatigue	
	$\beta^a$	OR <sup>b</sup>	$\beta^a$	OR <sup>b</sup>	$\beta^a$	OR <sup>b</sup>	$\beta^a$	OR <sup>b</sup>	$\beta^a$	OR <sup>b</sup>	$\beta^a$	OR <sup>b</sup>	$\beta^a$	OR <sup>b</sup>
Intercept	-1.51	-	2.30	-	-0.19	-	0.36	-	-2.43	-	-0.74	-	-3.20	-
Age (in years)	0.01	1.01	0.00	1.00	0.01	1.01	0.03	1.03	0.02	1.02	0.01	1.01	0.01	1.01
Sex (male = 0; female = 1)	0.16	1.17	-0.14	0.87	0.08	1.08	0.24	1.27	0.09	1.09	-0.01	0.99	0.13	1.13
Education level (high = 0)														
Mid (=1)	0.13	1.14	0.11	1.12	0.10	1.11	-0.03	0.97	0.13	1.14	0.03	1.03	-0.07	0.93
Low (=1)	0.00	1.00	0.14	1.15	-0.02	0.98	-0.11	0.90	-0.03	0.97	-0.10	0.90	-0.26	0.77
Number of comorbidities (none = 0)														
1 (=1)	-0.08	0.92	0.14	1.15	0.04	1.05	0.10	1.11	0.04	1.04	0.12	1.12	0.08	1.08
$\geq 2$ (=1)	0.06	1.06	-0.08	0.92	0.17	1.18	0.28	1.32	0.19	1.21	0.25	1.28	0.18	1.20
Time since diagnosis (in months)	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
Stoma (no = 0; yes = 1)	-0.10	0.90	0.00	1.00	0.08	1.08	-0.01	0.99	0.24	1.27	0.15	1.16	-0.13	0.88
Body mass index (in kg/m <sup>2</sup> )	0.01	1.01	-0.01	0.99	-0.04	0.96	0.01	1.01	-0.01	0.99	0.00	1.00	-0.01	0.99
Physical activity adherence (no = 0; yes = 1)	-0.21	0.81	0.02	1.02	-0.04	0.96	-0.36	0.70	-0.27	0.77	-0.13	0.88	-0.28	0.75
Anxiety score	0.01	1.01	-0.01	0.99	0.10	1.11	0.02	1.02	0.03	1.03	-0.01	0.99	0.03	1.03
Depression score	0.05	1.05	0.06	1.06	0.02	1.02	0.02	1.02	0.02	1.02	0.04	1.04	0.03	1.03
Fatigue score	0.01	1.01	0.01	1.01	0.01	1.01	0.01	1.01	0.01	1.01	0.01	1.01	0.03	1.03
Baseline HRQoL score	-0.03	0.97	-0.05	0.95	-0.03	0.97	-0.06	0.95	-0.02	0.98	-0.02	0.98	-	-
Chemotherapy (no = 0; yes = 1)	0.15	1.16	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	-0.18	0.83	0.00	1.00
Radiotherapy (no = 0; yes = 1)	0.16	1.17	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
Smoking (no = 0; yes = 1)	0.21	1.23	0.00	1.00	0.30	1.35	0.37	1.45	0.00	1.00	0.29	1.34	0.51	1.66

<sup>a</sup> Regression coefficients are the ln (odds) of change in the outcome, and can be used to calculate the probability of having low health-related quality of life =  $1/(1 + \exp[-\text{Linear predictor}])$ ; Linear predictor = intercept + sum of (regression coefficient \* predictors); for categorical and dichotomous predictors the reference category was always coded as 0.

<sup>b</sup> Odds ratios (OR) are shown to give an estimation of the strength of each predictor, but no confidence intervals or error measures could be calculated after recalibration of the slopes of the models, i.e., the shrinkage of regression coefficients. Note that the magnitude of the OR depends on the scale on which the predictor is measured and that the OR should not be causally interpreted.

expected [7]. We indeed observed that CRC survivors at baseline in the cohorts used for the current analyses reported somewhat lower HRQoL and functioning scores and higher fatigue scores on average than in the development cohort (Supplementary Table 2). This can likely be explained by the fact that having recently been diagnosed with and treated for cancer has a larger impact on an individual's physical and psychosocial well-being on the short-term as compared to the long-term. In the initial period after diagnosis and treatment, CRC survivors need time to mentally process and cope with the cancer experience as a major life event (e.g., fears of cancer recurrence and death) and to adjust to changes in their lives as a result of the cancer and treatment (e.g., living with a stoma and dealing with treatment complications). Although the prevalence of low HRQoL at follow-up was lower in all HRQoL domains in the current study than in the development study (Supplementary Table 2), the percentages of survivors showing a clinically relevant deterioration of HRQoL scores between baseline and follow-up were higher in the current study than the development study. This would suggest that the early post-treatment period within 6 month

after CRC diagnosis is a relevant time frame in the survivorship trajectory for the application of the risk prediction models to select high-risk survivors eligible for interventions aimed at HRQoL promotion. This time frame is considered a teachable moment during which survivors are open for interventions, providing a window of opportunity for taking tailor-made preventive actions.

The original prediction models included biopsychosocial predictors that had been selected based on evidence from studies showing associations with HRQoL, as summarized in our previously published systematic literature review [16]. For the updated models, we selected predictors for which the evidence in the systematic review was strongest, and that at the same time are relatively easy to assess in clinical practice (e.g., no long questionnaires or burdensome tests required). Moreover, the models have the advantage that they contain several modifiable predictors (5 out of 15), including lifestyle and psychosocial factors which have been shown to affect HRQoL in previous intervention studies [17–24]. It must be noted, however, that causal interpretations of relations between predictors and outcomes in risk prediction models should be made with great

caution, as the interpretation of such models should focus on the predictor set as a whole instead of focusing on isolated predictors. Prediction models are not meant to help unravel which predictors actually cause an outcome, such as low HRQoL.

Information on potentially effective interventions for improving HRQoL in CRC survivors is provided by several systematic reviews. Multiple studies have reported increased HRQoL after physical, psychological, or behavioral interventions in cancer survivors, including CRC [17–24]. Most evidence is available for exercise interventions, indicating that increasing post-treatment physical activity levels can improve HRQoL and functioning domains and decrease fatigue complaints. There is also evidence for beneficial effects of psychosocial interventions, such as cognitive behavioral therapy and psychoeducational or mindfulness-based interventions, on HRQoL and mental functioning domains. Nevertheless, some studies found less consistent effects of psychosocial interventions [46] or reported that not all domains of HRQoL were improved after exercise [18]. Future studies should therefore investigate whether the application of the prediction models for HRQoL can help to make a better selection of CRC survivors for specific interventions. For that purpose, so-called clinical impact studies are needed to evaluate whether application of the prediction models in oncology practice facilitates decision-making regarding appropriate interventions, and thereby contributes to improving HRQoL outcomes in high-risk CRC survivors. In addition, to promote implementation of the prediction models in practice, they should ideally be transformed into, for example, an online risk calculator or nomogram for clinicians to use for estimating their patient's 1-year risks of low HRQoL in the seven domains. Such tools are likely to be very relevant as they provide individual risk profiles that may guide subsequent risk-stratified follow-up care and that could also be used for doctor-patient communication about future risks and (modifiable) predictors. Application of such tools, which includes the assessment of predictors as input for the risk estimation, is becoming more feasible in practice nowadays due to enhanced digitalization of patient records within hospitals and increased attention for patient-reported outcome measures (such as HRQoL scales) as part of patient care.

A strength of the current study was its large sample size. Regardless of the smaller number of persons with low HRQoL at follow-up in the current study as compared to the development study, statistical power was adequate for the external validation according to the recommended minimum of 100 events and 100 nonevents for all HRQoL outcomes [44,45]. Another strength of the study is that we succeeded in downsizing the prediction models without losing predictive performance, so that a single set of predictors can be used for accurate estimation of 1-year risks of low HRQoL in multiple domains. Some limitations should also be mentioned. First, some predictors of the original

models were not or only partially available in the external validation cohorts (i.e., nine in the COLON and two in the EnCoRe cohort). As recommended based on results from simulation studies [38,39], we imputed missing predictors by adding the data from the cohort that was used for model development to the external validation datasets during the multiple imputation procedure, accounting for differences in time since diagnosis. Even though the missing predictors were eventually not included in the updated models, we cannot exclude the possibility that this procedure might have introduced bias in the external validation part of the analyses. A second potential limitation was that it was not known if and how many participants of the COLON study still received treatment at the time point that was used as study baseline for the current analyses (i.e., 6 months postdiagnosis). Some COLON study participants could thus still have been receiving treatment (chemotherapy) at that time, which might have influenced HRQoL scores at study baseline. Finally, the prediction models were developed and validated in Dutch populations of CRC survivors. This may limit the generalizability and applicability of our findings to CRC survivor populations in other countries. External validation in populations other than Dutch CRC survivors is to be recommended.

## 5. Conclusion

In CRC survivors within 6 months after diagnosis, we have externally validated and updated risk prediction models for estimating the 1-year risk of low HRQoL in multiple domains. To the best of our knowledge, these are the first validated models for predicting future HRQoL in the specific population of CRC survivors. Based on a parsimonious set of evidence-based biopsychosocial predictors, the performance of the models to provide accurate risk predictions and to identify high-risk CRC survivors was satisfactory. The models could be used to develop risk prediction tools that are applicable in oncology practice for identifying individual CRC survivors who have an increased risk of experiencing low HRQoL in the future and who may benefit from interventions aimed at improving or safeguarding their HRQoL. Clinical impact studies are warranted in order to evaluate the added value of applying the models in practice before the models can be implemented.

## CRedit authorship contribution statement

MB, MW, SvK, and DR have contributed to the conception and design of the work. DR, SvK, LS, and MB have done the analyses and interpretation of the data and results. All authors have contributed to the writing of the manuscript and have approved the final version.

## Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2022.09.019>.

## References

- [1] El-Shami K, Oeffinger KC, Erb NL, Willis A, Bretsch JK, Pratt-Chapman ML, et al. American cancer society colorectal cancer survivorship care guidelines. *CA Cancer J Clin* 2015;65(6):428–55.
- [2] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68(6):394–424.
- [3] Shapiro CL. Cancer survivorship. *N Engl J Med* 2018;379:2438–50.
- [4] Marventano S, Forjaz M, Grosso G, Mistretta A, Giorgianni G, Platania A, et al. Health related quality of life in colorectal cancer patients: state of the art. *BMC Surg* 2013;13(Suppl 2):S15.
- [5] Jansen L, Koch L, Brenner H, Arndt V. Quality of life among long-term (>=5 years) colorectal cancer survivors—systematic review. *Eur J Cancer* 2010;46(16):2879–88.
- [6] Jefford M, Ward AC, Lisy K, Lacey K, Emery JD, Glaser AW, et al. Patient-reported outcomes in cancer survivors: a population-wide cross-sectional study. *Support Care Cancer* 2017;25:3171–9.
- [7] Cabilan CJ, Hines S. The short-term impact of colorectal cancer treatment on physical activity, functional status and quality of life: a systematic review. *JBI Database Syst Rev Implement Rep* 2017;15(2):517–66.
- [8] Kawai K, Sunami E, Yamaguchi H, Ishihara S, Kazama S, Nozawa H, et al. Nomograms for colorectal cancer: a systematic review. *World J Gastroenterol* 2015;21:11877–86.
- [9] Hippisley-Cox J, Coupland C. Development and validation of risk prediction equations to estimate survival in patients with colorectal cancer: cohort study. *BMJ* 2017;357:j2497.
- [10] Engelhardt EG, Revesz D, Tamminga HJ, Punt CJA, Koopman M, Onwuteaka-Philipsen BD, et al. Clinical usefulness of tools to support decision-making for palliative treatment of metastatic colorectal cancer: a systematic review. *Clin Colorectal Cancer* 2018;17(1):e1–12.
- [11] Sales PM, Carvalho AF, McIntyre RS, Pavlidis N, Hyphantis TN. Psychosocial predictors of health outcomes in colorectal cancer: a comprehensive review. *Cancer Treat Rev* 2014;40(6):800–9.
- [12] Glavic Z, Galic S, Krip M. Quality of life and personality traits in patients with colorectal cancer. *Psychiatria Danubina* 2014;26(2):172–80.
- [13] Gray NM, Hall SJ, Browne S, Macleod U, Mitchell E, Lee AJ, et al. Modifiable and fixed factors predicting quality of life in people with colorectal cancer. *Br J Cancer* 2011;104:1697–703.
- [14] Revesz D, van Kuijk SMJ, Mols F, van Duijnhoven FJB, Winkels RM, Hoofs H, et al. Development and internal validation of prediction models for colorectal cancer survivors to estimate the 1-year risk of low health-related quality of life in multiple domains. *BMC Med Inform Decis Mak* 2020;20:54.
- [15] van de Poll-Franse LV, Horevoorts N, van Eenbergen M, Denollet J, Roukema JA, Aaronson NK, et al. The Patient Reported Outcomes Following Initial Treatment and Long Term Evaluation of Survivorship registry: scope, rationale and design of an infrastructure for the study of physical and psychosocial outcomes in cancer survivorship cohorts. *Eur J Cancer* 2011;47(14):2188–94.
- [16] Bours MJ, van der Linden BW, Winkels RM, van Duijnhoven FJ, Mols F, van Roekel EH, et al. Candidate predictors of health-related quality of life of colorectal cancer survivors: a systematic review. *Oncologist* 2016;21(4):433–52.
- [17] Buffart LM, Thong MS, Schep G, Chinapaw MJ, Brug J, van de Poll-Franse LV. Self-reported physical activity: its correlates and relationship with health-related quality of life in a large cohort of colorectal cancer survivors. *PLoS One* 2012;7:e36164.
- [18] Mishra SI, Scherer RW, Geigle PM, Berlanstein DR, Topaloglu O, Gotay CC, et al. Exercise interventions on health-related quality of life for cancer survivors. *Cochrane Database Syst Rev* 2012;2012: Cd007566.
- [19] Moug SJ, Bryce A, Mutrie N, Anderson AS. Lifestyle interventions are feasible in patients with colorectal cancer with potential short-term health benefits: a systematic review. *Int J Colorectal Dis* 2017;32(6):765–75.
- [20] Duncan M, Moschopoulou E, Herrington E, Deane J, Roylance R, Jones L, et al. Review of systematic reviews of non-pharmacological interventions to improve quality of life in cancer survivors. *BMJ Open* 2017;7(11):e015860.
- [21] Mishra SI, Scherer RW, Snyder C, Geigle P, Gotay C. Are exercise programs effective for improving health-related quality of life among cancer survivors? A systematic review and meta-analysis. *Oncol Nurs Forum* 2014;41(6):E326–42.
- [22] Mosher CE, Winger JG, Given BA, Shahda S, Helft PR. A systematic review of psychosocial interventions for colorectal cancer patients. *Support Care Cancer* 2017;25(7):2349–62.
- [23] Son H, Son YJ, Kim H, Lee Y. Effect of psychosocial interventions on the quality of life of patients with colorectal cancer: a systematic review and meta-analysis. *Health Qual Life Outcomes* 2018;16(1):119.
- [24] Turner RR, Steed L, Quirk H, Greasley RU, Saxton JM, Taylor SJ, et al. Interventions for promoting habitual exercise in people living with and beyond cancer. *Cochrane Database Syst Rev* 2018;9: CD010192.
- [25] van Roekel EH, Bours MJ, de Brouwer CP, Ten Napel H, Sanduleanu S, Beets GL, et al. The applicability of the international classification of functioning, disability, and health to study lifestyle and quality of life of colorectal cancer survivors. *Cancer Epidemiol Biomarkers Prev* 2014;23(7):1394–405.
- [26] Winkels RM, Heine-Broring RC, van Zutphen M, van Harten-Gerritsen S, Kok DE, van Duijnhoven FJ, et al. The COLON study: colorectal cancer: Longitudinal, Observational study on Nutritional and lifestyle factors that may influence colorectal tumour recurrence, survival and quality of life. *BMC Cancer* 2014;14:374.
- [27] Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;85:365–76.
- [28] Cocks K, King MT, Velikova G, de Castro G Jr, Martyn St-James M, Fayers PM, et al. Evidence-based guidelines for interpreting change scores for the European Organisation for the research and treatment of cancer quality of life questionnaire Core 30. *Eur J Cancer* 2012;48(11):1713–21.
- [29] Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Rheum* 2003;49:156–63.
- [30] Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. *J Clin Epidemiol* 2003;56:1163–9.
- [31] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatrica Scand* 1983;67(6):361–70.
- [32] World Cancer Research Fund/American Institute for Cancer Research. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Washington, DC: American Institute for Cancer Research; 2007.
- [33] Denollet J. DS14: standard assessment of negative affectivity, social inhibition, and Type D personality. *Psychosom Med* 2005;67(1):89–97.
- [34] Whistance RN, Conroy T, Chie W, Costantini A, Sezer O, Koller M, et al. Clinical and psychometric validation of the EORTC QLQ-CR29

- questionnaire module to assess health-related quality of life in patients with colorectal cancer. *Eur J Cancer* 2009;45(17):3017–26.
- [35] Winkleby MA, Jatulis DE, Frank E, Fortmann SP. Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *Am J Public Health* 1992; 82:816–20.
- [36] Darin-Mattsson A, Fors S, Kareholt I. Different indicators of socioeconomic status and their relative importance as determinants of health in old age. *Int J Equity Health* 2017;16(1):173.
- [37] van Buuren S. Package ‘mice’. 2017. Available at <https://cran.r-project.org/web/packages/mice/index.html>. Accessed July 1, 2017.
- [38] Held U, Kessels A, García Aymerich J, Basagana X, Ter Riet G, Moons KG, et al. Methods for handling missing variables in risk prediction models. *Am J Epidemiol* 2016;184:545–51.
- [39] Janssen KJ, Vergouwe Y, Donders AR, Harrell FE Jr, Chen Q, Grobbee DE, et al. Dealing with missing predictor values when applying clinical prediction models. *Clin Chem* 2009;55:994–1001.
- [40] Harrell FE Jr. Package ‘rms’. 2017. Available at <https://cran.r-project.org/web/packages/rms/rms.pdf>. Accessed July 1, 2017.
- [41] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36.
- [42] Hosmer DW, Stanley L. A goodness-of-fit test for the multiple logistic regression model. *Communications in statistics. Commun Stat* 1980;A10:1043–69.
- [43] Brier GW. Verification of forecasts expressed in terms of probability. *Monthly Weather Rev* 1950;78:1.
- [44] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual Prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1–73.
- [45] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual Prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg* 2015; 102(3):148–58.
- [46] Galway K, Black A, Cantwell M, Cardwell CR, Mills M, Donnelly M. Psychosocial interventions to improve quality of life and emotional wellbeing for recently diagnosed cancer patients. *Cochrane Database Syst Rev* 2012;11:Cd007064.