9

Heterogeneity in colorectal cancer symptom presentation general practice: A longitudinal observational study

Trynke Hoekstra
& Christel van Dijk, Sara Dorsman, Robert Verheij, Jos Twisk, François Schellevis
A revision has been submitted

Abstract

Background: The presentation of colorectal cancer (CRC) in general practice is complex and heterogeneous. Patients present with many non-specific symptoms that are also common in patients without CRC. This poses difficulties for general practitioners to quickly and adequately diagnose patients at high risk for CRC. This paper aims to support GPs in identifying patients at higher risk for CRC by employing an evaluation of symptom presentation in general practice in two ways. The first perspective aims to predict CRC-diagnosis as best as possible on the basis of the presence or absence of a range of symptoms, not taking into account the heterogeneity in symptom presentation. The second perspective aims to identify distinct subgroups of patients characterised by one symptom or various combinations of symptoms and is more focused on the heterogeneous symptom-presentation.

Methods: Data from a longitudinal observational study with 5 years' follow-up (LINH; Netherlands Information Network of General Practice) of 63,706 patients aged ≥40 years. The outcome measure was CRC-diagnosis over time. The presence/absence of 13 symptoms for CRC was included.

Two sets of analyses were conducted: a "traditional" analysis, namely a prediction model using all 13 symptoms, and a contemporary analysis aiming to reveal distinct subgroups of patients characterised by one symptom or (various) combinations of symptoms. These distinct subgroups of patients were identified by latent class analysis. To predict CRC-diagnosis over time, Cox regression analyses were conducted.

Results: The final prediction model included constipation, iron deficiency and rectal pain. The second set of analyses disclosed two distinct symptom profiles, where one group was predominantly symptom-free and another had a higher probability of reporting constipation, iron deficiency, weakness and/or abdominal pain. The latter subgroup, however, was not at increased risk of CRC-diagnosis (HR 1.12 (0.67-1.90)).

Conclusions: The extensive analyses presented in this paper did not provide additional clues for identifying patients at higher risk for CRC. The complex presentation of CRC in general practice therefore remains an urgent research topic.

Background

Identifying colorectal cancer (CRC) in general practice is a major challenge. The patient population is heterogeneous and presents with many non-specific symptoms which are also common in patients without CRC. These symptoms range from changes in bowel habits, anaemia or rectal bleeding to weight loss and weakness. Also, CRC-patients can be symptom-free for over 10 years [233–235] making it even more difficult for GPs to quickly identify patients at high risk for CRC. For these reasons, screening programmes are in place in several countries. Such programmes are relevant because CRC is one of the most common types of cancer in the western world with over 12,000 new cases in the Netherlands in 2009 [236] and CRC is well treatable when diagnosed early [237]. Unfortunately screening programs are operational in only a few countries and in most programmes participation rates are low [238, 239], so providing GPs with guidance in identifying patients at high risk for CRC is crucial. Yet, symptoms are vague or not present and referral guidelines focus mainly on a genetic predisposition (including a positive family history) only [240] meaning that GPs have relatively little to rely on.

The current literature certainly helps GPs in decision making, but does not provide them fully with the answer to the question which patient has a possible elevated risk of CRC based on his or his symptom presentation and thus needs to be referred for further diagnostics.

Studies investigating thoroughly the symptom presentation of CRC in general practice are relatively rare. These studies, for example, look only at one symptom or a fixed combination of symptoms [241–244] at one time point predicting CRC at another time point, i.e. not taking into account the heterogeneous patient population. What is therefore urgently needed is an evaluation of symptom-presentation in general practice and its relation to subsequent CRC-diagnosis, taking into account the acknowledged individual patient differences. This is also recently acknowledged by a systematic review [245], showing that diagnostic studies in CRC in primary care are scarce. The authors urge the need for more investigation of CRC-symptom presentation in general practice. Moreover, another systematic review [246] indicated the relatively low predictive value of symptoms for CRC, recommending future studies to include a broader range of symptoms rather than assessing selected symptoms in isolation. Such studies are currently limited, yet crucial to take this field of research a step further.

Therefore, this paper will employ an evaluation of symptom presentation and CRC-diagnosis in general practice, taking into account individual patient differences, by using longitudinal observational data over five years.
The aim of this paper is not to provide GPs with a perfect model based on the available patient information in general practice, but to contribute to a better identification of patients at increased risk of CRC based on their symptom presentation
First, we will develop a prediction model using all available symptoms to identify patients with an increased risk of CRC-diagnosis over time. This perspective aims to predict CRC-diagnosis as best as possible on the basis of the presence or absence of symptoms, not necessarily taking into account the heterogeneity in symptom presentation. Second, by making use of all possible symptoms for CRC without preselected combinations of symptoms, we identify distinct subgroups of patients characterised by one symptom *or* various combinations of symptoms. Assessing the risk of CRC-diagnosis over time of these subgroups could reveal characteristics of patients specifically at risk for CRC-diagnosis. Furthermore, this perspective is more focused on GP-patients themselves by looking at the heterogeneity of symptoms presented in general practice. Comparing two ways of looking at this complex issue could provide GPs with important information regarding the question which patient to refer for further diagnostics.

Methods

Study design

Data were derived from the Netherlands Information Network of General Practice (LINH) for 2002-2006. The LINH database holds longitudinal data on morbidity, mortality, drug prescribing and referrals of about 340,000 individuals listed in general practice. Anonymised data on all patient contacts, including diagnoses, referrals and drug prescriptions are derived from the routinely kept electronic medical records in the participating practices. Diagnoses are coded using the ICPC (International Classification of Primary Care [225]). The network consists of a dynamic pool of practices, with a limited annual turnover. All data are collected and handled according to the data protection guidelines of the Dutch Data Protection authority.

Participating practices
The following inclusion criteria for the use of practice data were applied: a) GPs assigned diagnostic ICPC-codes to 50% or more of their consultations, and b) consultations and drug prescriptions were recorded adequately throughout the year (at least 10% of all consultations and 10% of all drug prescription had to be recorded every three months). In total, data from 37 practices were included in the current study, with an average assignment of diagnostic ICPC-codes to consultations of over 85%. These practices had valid data for the entire period of 2002-2006 and were representative of the Dutch GP-practices with respect to degree of urbanisation and region, but not in respect to practice type (single handed, duo, group or health centre).

Participating patients
Only patients who were aged 40 or over in 2002 were included, since CRC is not very common in younger ages. Patients diagnosed with any other form of cancer (ICPC-codes A79, B72-B74, D74, D76, D77, N74, R84, R85, S77, T71, U75-U77, W72, X75-X77, Y77 or Y78) were excluded from the analyses. Also, patients whose electronic record included the diagnosis of CRC in 2002 or before were excluded. In total, data from 63,706 patients were included in the current study.

Measurements - colorectal cancer
The diagnosis of colorectal cancer was determined by the ICPC code D75; malignant neoplasm colon/rectum. Only newly diagnosed cases of colorectal cancer in 2003-2006 were taken into account, to be certain that symptoms preceded CRC diagnosis.

Measurements - symptoms
Thirteen symptoms (present/absent) presented during consultations in 2002 were included in the analyses based on available cancer literature [241–244, 247–254]. These included weakness/tiredness (ICPC code A04), feeling ill (A05), iron deficiency anaemia (B80), abdominal pain (D01 or D06), rectal/anal pain (D04), perianal itching (D05), diarrhoea (D11), constipation (D12 - a "protective" symptom [248, 253]), meleana (D15), rectal bleeding (D16), change in bowel habits (D18), bloating (D24 or D25) and weight loss (T08).

Statistical Analysis

Two sets of analyses were conducted. First, a full prediction model was constructed with all available symptoms (independent variables) in relation to time to CRC-diagnosis (dependent variable). A Cox regression model was conducted using a backward selection. One by one symptoms with the lowest predictive value were deleted from the model using a threshold of $P<0.15$ to accommodate the low number of cases. Age (as a categorical variable; quartiles) and gender were forced in the model irrespective of their *P*-values.

Second, we used Latent Class Analyses (LCA, [5, 6, 18, 42]) to identify distinct classes of patients with specific (combinations of) symptoms. LCA are types of cluster analyses used to group patients into *k* number of unique (otherwise unobserved) categories, where, within each category patients are most similar to each other regarding their CRC-symptom profile, and between the categories patients are most different. The aim of the technique is to find the optimal number of categories to represent the data best. To find the optimal number of categories, a 1-4 class solution was modelled and output was assessed and compared. To determine the "best" model, model fit parameters such as the Bayesian Information Criterion (BIC; where a lower BIC implies better fit [66]) were used. Also, we assessed the usefulness and clinical interpretation of each solution. This was done by assessing the solutions based on the number of people in each class (hereby rejecting solutions with relatively small sample sizes), and by assessing the symptom profiles based on the CRC-literature (hereby rejecting solutions with combinations of symptoms that do not make theoretical/clinical sense).

This provided us with a categorical variable, representing the *k* number of classes. Each patient belongs to one class only. Subsequently, this categorical variable was used in the Cox regression analyses. These models were conducted to analyse CRC-diagnosis over time. Also this model was adjusted for age (as a categorical variable; quartiles) and gender.

All analyses were run using the Stata 11.0 and Mplus 6.12 [34] statistical software programs.

Results

In total, 63,706 patients were included in the study, of which 50.9% were female. Mean age (standard deviation) of the participants was 54.4 (12.7) at baseline.

In **table 20** characteristics of the study population are presented, including the prevalence of the thirteen symptoms in 2002 and the incidence of colorectal

cancer at subsequent time points. The prevalence rates of the symptoms vary considerably (meleana having the lowest prevalence with just over 1 person in every 1000 and abdominal pain having the highest prevalence with over 50 persons in every 1000 persons).

Table 20 Prevalence of symptoms in 2002 separated by colorectal cancer diagnosis

	Diagnosed with CRC during follow-up	**Never diagnosed with CRC**
Meleana:	3.84 per 1000	1.06 per 1000
Bloating:	0.00 per 1000	1.62 per 1000
Rectal pain:	6.97 per 1000	1.95 per 1000
Feeling ill:	3.84 per 1000	1.65 per 1000
Change in bowel habits:	0.00 per 1000	2.53 per 1000
Weight loss:	6.97 per 1000	3.47 per 1000
Rectal bleeding:	13.94 per 1000	6.31 per 1000
Perianal itching:	6.97 per 1000	7.50 per 1000
Diarrhoea:	34.84 per 1000	15.10 per 1000
Iron deficiency anaemia:	45.30 per 1000	17.91 per 1000
Constipation:	38.33 per 1000	34.64 per 1000
Weakness:	66.20 per 1000	41.48 per 1000
Abdominal pain:	80.14 per 1000	52.58 per 1000

Colorectal cancer prevalence (cumulative)

CRC 2003: 0.83 per 1000
CRC 2004: 1.82 per 1000
CRC 2005: 3.14 per 1000
CRC 2006: 4.51 per 1000

Symptom based prediction model

The results of the prediction model are shown in **table 21**. After backward selection three symptoms (constipation, iron deficiency anaemia and rectal pain) were identified as independent predictors for CRC-diagnosis over time. Constipation showed to have a protective effect for being diagnosed with CRC (HR: 0.63). Both iron deficiency anaemia and rectal pain showed to be a risk factor (HR 1.97 and 2.97 respectively).

Table 21 Final multivariate Cox regression model with predictors for time to CRC-diagnosis

	HR	**SE**	**95% CI**	***P*-value**
Constipation	0.63	0.19	0.34 to 1.15	0.13
Iron Deficiency Anaemia	1.97	0.56	1.12 to 3.45	0.02
Rectal pain	2.70	2.11	0.74 to 11.97	0.13

Symptom (patient) profiles

In **figure 8** the results of the latent class analyses can be seen. On the x-axis the thirteen symptoms are plotted against the (mean) prevalence of each symptom in each class. The best model was a 2-class solution where most patients were classified into a predominantly symptom free subgroup (N=61,726, 96.9%). Further, a smaller subgroup (N=1,980, 3.11%) was identified in which patients showed relatively high probabilities of certain symptoms (constipation, iron deficiency anaemia, weakness and abdominal pain). These symptoms clearly differentiate the two subgroups from each other.

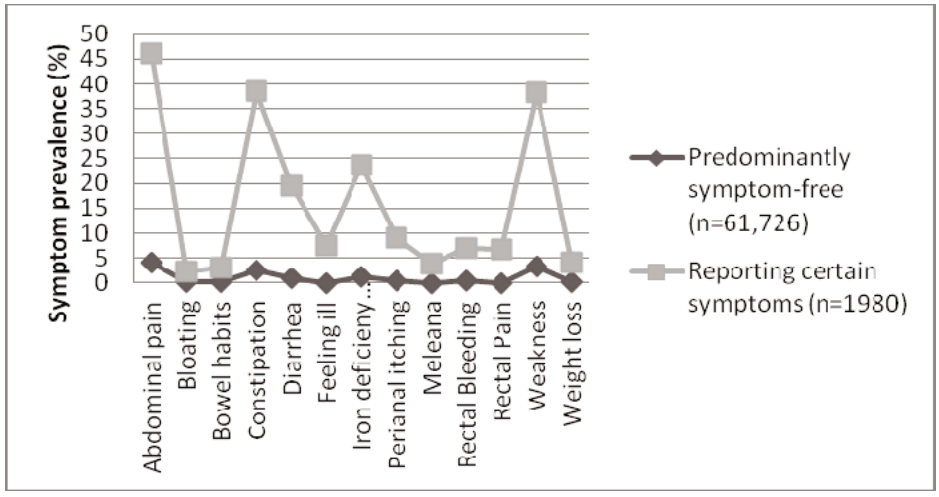

Figure 8 Two distinct symptom profiles in general practice

The results of the Cox regression models including the two classes are presented in **table 22**. The crude model shows that the smaller class with symptoms has a higher risk of CRC-diagnosis over time compared to the symptom-free class. This increased risk is mostly explained by age and gender differences between the two classes (model 2). Inspecting the characteristics of the classes showed for example that patients classified into the symptom-free class are younger (mean age 62.11, standard deviation 12.46) compared to the class with a higher prevalence of selected symptoms (mean age 70.16, standard deviation 15.79).

Table 22 Hazard ratios, 95% confidence intervals and corresponding *P*-values for the Cox regression models comparing time to CRC-diagnosis of the two classes

	HR	**SE**	**95% CI**	***P*-value**
Crude model				
Class with certain symptoms[1]	1.72	0.46	1.02 to 2.89	0.04
Adjusted model (gender, age)				
Class with certain symptoms[1]	1.12	0.30	0.67 to 1.90	0.66

[1] "Symptom-free" class set as reference

Integration of the results

The results of the two sets of analyses are difficult to compare *directly*. In the prediction model the predictive value of separate symptoms and the number of symptoms were tested. However, in the latent class analyses, distinct classes of patients with specific (combinations of) symptoms were constructed without taking into consideration effects of these separate symptoms on CRC-diagnosis over time. Therefore, these analyses show the clustering of symptoms common in general practice, and not the symptoms showing the highest risk of CRC-diagnosis over time. Nonetheless, the clustering of patients with specific symptoms can be *indirectly* compared to the results of the prediction model. If the symptoms of the prediction model are consistent with the symptoms within the clustering of patients, GPs can identify certain patients or groups of patients specifically at risk for CRC-diagnosis over time common in general practice. Also, a specific cluster of patient could show to have a high risk in itself on CRC-diagnosis. However, the results show no consistency in this respect.

Discussion

Identifying patients at high risk for colorectal cancer in primary care is a major challenge for GPs. The current study aimed to contribute to a better identification by analysing its heterogeneous presentation in general practice in two different ways: 1) by utilising all known symptoms related to CRC, we developed a prognostic model predicting CRC-diagnosis over time, and 2) by attempting to reveal possible subgroups of patients with comparable symptom profiles with distinct CRC-risk. Constipation, iron deficiency anaemia and rectal pain predicted time to CRC-diagnosis in the final prediction model, where constipation reduced the risk of CRC-diagnosis over time. The second set of analyses disclosed two distinct symptom profiles, where one group was predominantly symptom-free and another had a substantially higher probability of reporting constipation, iron deficiency, weakness and/or abdominal pain. This subgroup then showed a significantly higher CRC-diagnosis risk over time compared to the subgroup with a relatively low probability of any symptom. This higher risk, however, was mainly explained by age and gender differences between the two classes.

Strengths and weaknesses of the study

We used an observational cohort of GP-patients representative of the general older (40^{+} years) GP-patient population. This study sample, with over 60,000 patients, reflects what GPs actually see in their practice, taking into account a broad range of possible symptoms for CRC. Together with a follow-up period of five years after presenting symptoms this allows for a very thorough investigation of our research questions.

However, our study has some limitations. In the CRC literature, importance is often placed on the vagueness of the symptoms or the long symptom free period of the disease. In this light, the *duration* [250] of the symptoms might play a central role in many scientific studies and also in the medical examinations of the patients in the clinic. Unfortunately, this information was not available; we were only able to assess whether or not (yes/no) a patient reported a symptom. However, studies looking at duration of symptoms in relation to CRC-diagnosis seem inconsistent [250, 255]. Also, because CRC-related symptoms are often vague, common symptoms, patients who visit their GP with these symptoms are most likely the ones suffering from these symptoms long enough to consult their GP on it. The Dutch GP-guidelines recommending whether a patient should be referred for further testing include a genetic predisposition,

or a positive family history of polyps [236]. This information is not recorded systematically and routinely by GPs in the Netherlands, so we could not include this variable as a potential confounder in our analyses. However, research [256] has shown that self-reported family history of polyps (or being genetically predisposed) can be inaccurate, showing that these data should be interpreted with caution if available. This does not mean that the genetic predisposition, or family history is meaningless and this information should be used in the decision making process.

Meaning of the study

Existing research consistently shows that especially iron deficiency anaemia is an alarm symptom for CRC [244, 245]. This was confirmed in our results. Further, studies are fairly inconsistent on other possible alarm symptoms for CRC [234, 242, 249, 250]. A change in bowel habits is sometimes mentioned, but also rectal problems, weakness and weight loss [246] (these last two symptoms are often present in cancer in general). Partly, our results are consistent with these findings; rectal problems (i.e. rectal pain) showed to have a predictive value for CRC-diagnosis. Our second set of analyses investigated the possible clustering of symptoms in GP-patients. Patients in the symptom cluster showed a more frequent reporting of abdominal pain, constipation, iron deficiency anaemia and weakness. The clustering did not result in a higher CRC-diagnosis risk, which can be explained by the lack of many alarm symptoms as found in the first set of the analysis, and the fact that constipation has been shown to be a protective symptom for CRC-diagnosis [248]. However, it does show that patient symptoms do cluster, but their relevance remains unclear.

Overall, the results reveal that patients do not show a clear presentation of alarm symptoms for CRC in general practice, making it still difficult for GPs to decide which patient to refer for further diagnostics based on their symptom presentation in general practice. Only the presence of iron deficiency anaemia clearly distinguishes CRC-patients, as well as older age, showing that GPs still have very little to rely on.

Based on the vague presentation of CRC in general practice, various countries in Europe and elsewhere have been, or are in the middle of implementing screening programmes for CRC to enhance early diagnosis and treatment [257]. However, compared to other cancer screening programmes (for example breast cancer or cervical cancer) participation rates remain low in most programmes ([238, 239, 258]. These low participation rates are widely acknowledged in

the literature [238, 239, 258] and by health (policy) workers also [257, 259]. However, simply by acknowledging the lack of success of CRC screening programmes does not make the need for early detection less important. GPs currently still need information about which patient to refer for further testing. A referral of every patient with CRC-related symptoms is not desirable as this would place an extensive burden on secondary health care and would leave patients with possible unnecessary feelings of stress and worry because of the referral to secondary care. It remains essential to explore other ways to help GPs and their patients. Analysing symptom presentation in general practice in multiple ways, with extensive, longitudinal data has not (yet) provided the answer for GPs.

Unanswered questions and future research: concluding remarks

Although our study adds important information about CRC presentation in general practice, possible CRC-patients will still be missed when going by our results alone; not every CRC-patient (either with- or without a genetic predisposition) reports to GPs with identical symptoms. Presumably, here also the GP's gut feeling plays a part [260]. A recent study shows that, when it comes to cancer, ambiguous complaints like 'duration of symptoms', 'the appearance of a patient', or 'a patient that usually rarely visits the GP' can trigger the GP's gut feeling and stimulate referral to a specialist for further diagnostics [36]. For CRC, it would be worthwhile to investigate and specify certain symptoms, or patient characteristics in this respect.

The current study shows us that every CRC-patient is different, and possibly that this patient-heterogeneity cannot be unravelled fully (yet). Therefore, future research looking into raising the awareness of CRC-risk, both in at-risk individuals themselves and GPs is warranted. It is also crucial to further explore the possibilities of successful screening programmes [257, 261] as this might be the only way forward.

Acknowledgements

We would like to thank all staff at the practices of the Netherlands Information Network of General Practice for their cooperation.