Self-reported information on the diagnosis of colorectal cancer: reliable, but not necessarily valid

Brigid M. Lynch^a, Danny Youlden^a, Lin Fritschi^b, Beth Newman^c, Ken Pakenham^d, Barbara Leggett^e, Neville Owen^f, Joanne F. Aitken^a

^a Viertel Centre for Research in Cancer Control, The Cancer Council Queensland

^b Western Australian Institute for Medical Research

^c School of Public Health, Queensland University of Technology

^d School of Psychology, The University of Queensland

^e Department of Gastroenterology and Hepatology, Royal Brisbane and Women's Hospital

^f Cancer Prevention Research Centre, School of Population Health, The University of Queensland

Address for correspondence: Brigid M. Lynch

Viertel Centre for Research in Cancer Control

The Cancer Council Queensland

PO Box 201

Spring Hill 4006

Australia

Tel: +61 7 3258 2314

Fax: +61 7 3258 2310

Email: brigidlynch@qldcancer.com.au

Abstract

Objective: Self-report is commonly used in epidemiologic studies, however little data exists on the reliability and validity of this method for eliciting information related to the diagnosis of colorectal cancer. We examined the test-retest reliability and validity of colorectal cancer patient reporting on the process of their diagnosis.

Study design and setting: One hundred and sixteen participants completed two telephone interviews, one month apart, and 95 general practitioners completed a written questionnaire, to elicit information relating to key elements of the process of diagnosis of colorectal cancer.

Results: Acute symptoms such as rectal bleeding had higher reliability and validity than more general symptoms. Colonoscopy was the most accurately recalled diagnostic test. Recall of diagnosis date, and date of colonoscopy, had high test-retest reliability. There were considerable differences between dates of diagnostic tests given by participants and general practitioners, but there was no evidence of a bias in a particular direction. Accuracy of recall did not diminish as time from diagnosis increased.

Conclusion: This study confirms that self-reported symptoms, tests and dates in the colorectal cancer diagnostic pathway are generally reliable, however the validity of reported symptoms and tests can be moderate to poor.

Keywords: test-retest reliability; validity; epidemiologic methods; self-report; colorectal cancer; colonoscopy

Running title: Self-reported colorectal cancer diagnosis

Word count (abstract): 193

Word count (text): 3,540

1. Introduction

In the United States of America, Europe and most of the industrialised world, colorectal cancer is the third most common cancer and the second most common cause of cancer deaths (1-3). Risk increases progressively with age, and the incidence of colorectal cancer is expected to increase as worldwide trends in population ageing continue (1).

The process of diagnosis for colorectal cancer is often characterised by multiple symptoms and a number of diagnostic tests. Symptoms may range from the acute, such as rectal bleeding, to more general, constitutional complaints, such as lethargy and loss of appetite. Symptoms may not be apparent until late in the disease course, and as a result, colorectal cancer is often diagnosed at an advanced stage (4). Patients sometimes undergo a number of tests before a conclusive diagnosis is made.

Self-report is a common method of obtaining data pertaining to symptoms and diagnostic tests for epidemiologic studies, primarily because it is easy to implement and inexpensive (5-8). However, self-reported data may be limited by a number of factors, such as recall error, biases such as social desirability and acquiesce, and by poorly-constructed measures (7, 9-11).

There is a scarcity of data on the reliability of self-reported colorectal cancer symptoms. Malats and colleagues (12) compared hospital records with interview data to assess agreement on the initial symptom experienced by patients with cancer of the digestive tract (just over half had colorectal cancer). They found that the concordance between hospital records and interview data was 61% for type of first symptom, with only 46% agreement for the date of first symptom (within 30 days).

A number of studies have assessed the validity of self-reported testing for colorectal cancer (5, 13-16). However these studies focussed on colorectal cancer screening in the general population. There are no published data on the validity of self-reported diagnostic testing

among colorectal cancer patients. Most studies of the validity of self-reported screening tests found at least moderate agreement between self-report and other, more objective data (13-16). Colonoscopy and sigmoidoscopy were more accurately recalled than faecal occult blood testing; recall was generally better for more invasive tests (5, 14). A common finding was that participants tended to report having had screening tests more recently than had been the case.

We examined the test-retest reliability and the validity of patient self-report on key elements of the process of diagnosis of colorectal cancer.

2. Methods

2.1 Sample

One hundred and eighty-one adults were randomly selected from participants in a population-based, longitudinal study of the diagnosis of colorectal cancer and quality of life following diagnosis. All had been diagnosed with a first, primary diagnosis of colorectal cancer between 1 January 2003 and 31 December 2004, were aged between 20 and 80 years, and had been recruited through the Queensland Cancer Registry. During a telephone interview, these 181 potential participants were asked whether they would take part in a repeat interview, one month later; 148 agreed to be re-contacted and 116 were successfully re-interviewed. The University of Queensland's Behavioural and Social Science Ethical Review Committee approved the study's procedures.

2.2 Data Collection

Open-ended questions were asked during both telephone interviews to ascertain details of symptoms or reasons for visiting a doctor. Participants were also asked how long they had these symptoms before consulting their doctor, the date of the first doctor consultation, how many consultations they had prior to diagnosis, the type and date of tests performed by their doctor, and the date of diagnosis. Sociodemographic information was collected at the initial telephone interview only.

Participants also provided contact details for their general practitioner (GP) at the initial interview. GPs were mailed a one-page survey, asking about date of first presentation; reason for initial presentation, including symptoms; diagnostic tests performed and dates of those tests.

For both the patient interview and the GP survey, the non-mention of a symptom or test was assumed to correspond with a “no” response.

2.3 Statistical Analyses

2.3.1 Reported symptoms and diagnostic tests

A variety of statistical methods are available to test reliability and validity (17). To allow comparison of results within the study, we used the same measures for both reliability (comparing data from Interviews 1 and 2) and validity (comparing results from Interview 1 to information supplied by the participant’s GP). In order to present meaningful estimates of reliability and validity, analyses were only conducted for symptoms or tests that had been reported by at least 10% or more of participants. All analyses were conducted using SAS version 9.1 (SAS Institute, Carey, NC).

Two summary statistics have been calculated for each symptom or test. Positive specific agreement is a measure of the proportion of concordant positive results from the two datasets being compared, and is calculated using $\frac{2 \times n_{pa}}{(2 \times n_{pa}) + n_d}$, where n_{pa} = the number of observations with positive agreement [(yes,yes)] and n_d is the number of discordant observations [(yes,no) or (no,yes)] (18, 19). The kappa coefficient provides a measure of the overall agreement after correcting for chance, given by the general formula $\frac{A_o - A_e}{1 - A_e}$, where A_o = observed agreement and A_e = expected agreement (19, 20).

Positive specific agreement was used in preference to the uncorrected index of overall agreement, because the latter may have resulted in artificially high estimates of concurrence due to the open-ended nature of the questions regarding symptoms and tests: negative agreement [(no,no)] may occur due to an oversight by the respondent or GP, rather than reflecting the true absence of that symptom or test. All calculations for both positive specific agreement and kappa relating to the symptoms and diagnostic tests were based on dichotomous (yes/no) data.

There are no definitive standards for interpreting positive specific agreement or kappa. Table 1 summarises how the indices have generally been interpreted for this study, in line with recommendations made in the literature (17, 18, 21). Positive specific agreement has been expressed as a percentage, and kappa as a ratio between -1 and 1.

INSERT TABLE 1 ABOUT HERE

2.3.2 Date of diagnostic tests, date of diagnosis and date of first consultation

Comparisons between the reported dates of colonoscopy (Interview 1 versus either Interview 2 or the GP data) and date of diagnosis (Interview 1 versus Interview 2 only) were summarised by calculating the percentage of the dates that agreed exactly, agreed within one week, agreed within one week to one month, or differed by one month or more. The date of diagnosis was not included on the GP survey form, therefore no validity analysis was conducted for this data item. Comparisons of date of first consultation could only be made at a broader level (same month or one month or more earlier or later), as this was only recorded to the nearest month for a considerable proportion of respondents. Accuracy of recall of the above dates was further assessed by calculating weighted kappa coefficients. Weighted kappas are an extension of the

simple kappa coefficient using weights to quantify the relative difference between responses (22). Weights were derived using the Cicchetti-Allison method (23).

The difference in the dates was also plotted against the date given at Interview 1 to try to establish whether there was a relationship between the size of the difference and the length of time since the event was initially reported to have occurred. This approach is based on the Bland-Altman method (24), which is commonly used to examine the difference of two measurement techniques against their average, to identify any systematic bias. A regression line was also fitted to these plots to assess the significance of the relationship. A small number of extreme outlier values were removed from these analyses so that they did not overwhelm the results. For the purpose of this study, an extreme outlier was defined as a value $< Q1 - 3 \times IQR$ or $> Q3 + 3 \times IQR$, where $Q1$ = first quartile, $Q3$ = third quartile and IQR = interquartile range (25).

There were insufficient data to evaluate the reliability or validity of other diagnostic tests apart from colonoscopy.

2.3.3 Number of doctor consultations

To calculate the test-retest reliability of the reported number of visits to a doctor prior to diagnosis, the overall agreement and weighted kappa coefficients were used.

2.3.4 Time from diagnosis and accuracy of recall

The reliability and validity of each variable were correlated against time from diagnosis to first interview, using the Spearman method (22). This was to determine whether the accuracy of recall diminished as the time from the diagnostic process increased.

3. Results

The mean and median time between Interview 1 and Interview 2 was 34 days (range = 27 to 44 days). The sociodemographic and disease-related characteristics of the study participants are presented in Table 2.

INSERT TABLE 2 ABOUT HERE

3.1 *Reported Symptoms*

There was a moderate to high degree of reliability between Interview 1 and Interview 2 for more specific symptoms of colorectal cancer (Table 3). “Blood in stools”, “pain in lower abdomen” and “change in bowel habit” all had positive specific agreements between 78% to 90%, and kappa coefficients ranging from 0.72 to 0.85. “Lack of energy/tiredness”, a symptom common to a range of illnesses, only had a moderate level of reliability. There were insufficient data to do meaningful analyses of the reliability of other specific symptoms, such as nausea or vomiting, unexplained weight loss or loss of appetite.

Validity for reported symptoms was moderate for “blood in stools”, “pain in lower abdomen” and “change in bowel habit”, which all resulted in positive specific agreements of 59% to 66% and kappa values between 0.43 to 0.50 (see Table 4). For the less specific symptom of “lack of energy/tiredness”, both the positive specific agreement and kappa coefficient indicated a poor level of validity.

3.2 *Diagnostic tests*

Colonoscopy was the most commonly reported test, with 107 participants (92%) stating during either interview that they had a colonoscopy. This diagnostic test had the highest levels of

positive specific agreement (94%), however the value of the kappa coefficient (0.54) did not support the high level of reliability suggested by the former index.

Of the other diagnostic tests, faecal occult blood test, ultrasound, and X-ray all had moderate to high levels of reliability, with positive specific agreement ranging from 73%-78% and kappa coefficients between 0.70-0.76. The results also indicated moderate reliability for digital rectal examination, blood test and CT scan (see Table 3).

INSERT TABLE 3 ABOUT HERE

In terms of validity, colonoscopy again had by far the highest positive specific agreement (91%), but the kappa coefficient was on the borderline of poor/moderate (0.39), reflecting the correction for chance agreement. Digital rectal examination, faecal occult blood testing and blood test all showed moderate levels of validity based on both the positive specific agreement and kappa statistics, while validity was in the poor range for CT scan, ultrasound and X-ray.

INSERT TABLE 4 ABOUT HERE

3.3 Date of First GP Consultation

Participants had difficulty in recalling the precise date on which they first consulted their doctor about the symptom or other reason that led to colorectal cancer being diagnosed. Of the 115 eligible participants, 43 (37%) were unable to give an exact date for their initial consultation with a GP at Interview 1, and 50 (43%) were unable to nominate an exact date at Interview 2. Therefore, these data could only be matched to the nearest month. Even with this reduced

precision in matching the dates there was still considerable variation in recall, with almost a quarter of participants giving dates that were at least one month or more apart (8% earlier and 16% later than reported at Interview 2). However, based on the kappa coefficients for agreement within the same month, reliability for date of first GP consultation was moderate to high, with a weighted kappa of 0.85 (0.77–0.93).

The Bland-Altman plot (Figure 1) displays no clear pattern of dependency between the date nominated during Interview 1 and the difference in the dates reported between Interviews 1 and 2 for the majority of the data points. When the extreme outliers were removed (retaining 109 out of 115 data points), the slope of the resulting regression line was not statistically significant ($p=0.087$).

INSERT FIGURE 1 ABOUT HERE

When the date of first consultation reported at Interview 1 and the GP data were compared ($n=87$), there was agreement on the same month for only 43% of participants, with the rest of the cases recording dates that were at least one month apart. Twenty-three (26%) nominated a date that was one month or more earlier than the date from the GP, while 27 (31%) nominated a date that was one month or more later. The poor to moderate validity is also borne out in the weighted kappa (0.51, 95% CI = 0.38–0.64).

A greater amount of bias is evident in the Bland-Altman plot comparing the difference between Interview 1 and the GP data for date of first doctor consultation than in the comparison between the two interviews. Even after removing extreme outliers (retaining 83 out of 87 data points), the slope of the line remains highly statistically significant ($p<0.001$). Figure 2 suggests that if an earlier date (prior to January 2003) was nominated by the patient during Interview 1, this was

generally six to 12 months before the date reported by the GP. Conversely, there were several observations where the data reported in Interview 1 were more recent (between April 2003 to December 2003), but the date provided by the GP was considerably earlier than this.

INSERT FIGURE 2 ABOUT HERE

3.4 Date of diagnostic tests (colonoscopy only)

For reported dates of the diagnostic tests, the date of colonoscopy had high reliability, with 90% of eligible participants (n=91) reporting the same date at both interviews. Only one respondent reported a date that was more than one month earlier than reported at Interview 2, and only one participant reported a date that was more than one month later. This was further underlined by the weighted kappa coefficient for agreement within the same month and year of 0.98, 95% CI = 0.95–1.00.

There was more difference in the dates reported for colonoscopy between Interview 1 and the GP data than was observed between the two interviews. Exact agreement on date of colonoscopy fell to 46%, with 12% of the dates more than one week earlier for Interview 1 in relation to the GP data, and 17% at least one week later. The weighted kappa coefficient for testing validity for the same reported month and year for colonoscopy was 0.75 (95% CI = 0.60–0.90), which is still indicative of a moderate to high level of validity.

There was no obvious pattern in the Bland-Altman plots for date of colonoscopy when comparing Interview 1 against either Interview 2 or the data obtained from the respondent's GP (graphs not shown).

3.5 Diagnosis date

A comparison of diagnosis dates reported at Interview 1 and Interview 2 showed very good reliability, with 85% of respondents (n=116) in exact agreement between the two interviews. Only 3% of respondents reported dates that were more than 1 month different (either earlier or later) during Interview 1 compared to Interview 2, with a corresponding weighted kappa of 0.97 (95% CI = 0.93–1.00).

3.6 Number of doctor consultations

For the number of doctor consultations reliability was moderate, with observed agreement of 64% (n=116) between Interviews 1 and 2. However, a further 32% of respondents differed by only one visit in the response that they gave for Interview 1 in comparison to Interview 2, and the mean number of visits (2.8, 95% CI = 2.6 - 2.9) before colorectal cancer diagnosis was the same for both data collection points. The kappa coefficient also points to a moderate level of reliability, with an a weighted kappa of 0.57 (95% CI = 0.45–0.69).

3.7 Time from diagnosis and accuracy of recall

There were no significant associations between time from diagnosis to first interview and either test-retest reliability or validity, for any of the variables assessed in this study (data not shown).

4. Discussion

To our knowledge, this is the first study to assess the test-retest reliability and validity of self-reported cancer diagnostic tests. Such information is important because self-report will continue to be widely used in cancer epidemiology studies. Our findings suggest that the accuracy of self-reported diagnostic information is variable, depending on the type of information being elicited. Overall, we found the test-retest reliability to be moderate to high, and the validity moderate to poor. We did not find that accuracy of recall diminished as time from diagnosis increased.

In terms of symptoms, the more acute symptoms specific to colorectal cancer were found to have good reliability and moderate validity. However, this was not the case for less specific symptoms. This may be because patients only report their more acute symptoms to their GP, or it may be the case that GPs do not routinely record symptoms that they do not consider important. The limited time allocated for GP consultations may impact both on patients' tendency to report all symptoms and time of onset, and the comprehensiveness of GPs' records.

Colonoscopy was the most accurately recalled diagnostic test in terms of uncorrected agreement, however, the values of the kappa coefficient did not appear to support the high levels of reliability and validity suggested by the positive specific agreement. This apparent anomaly may be due to either the open-ended nature of the question and/or the high positive response in this case. These factors most likely contributed to the value obtained for the negative specific agreement (60% - moderate), and the imbalance between the positive and negative agreement would then impact on the calculation of kappa (19). Given the characteristics of the data, it is likely that the positive specific agreement provides a better indication of the reliability and validity of the colonoscopy responses than the kappa coefficient (23), although the "true" answer is probably somewhere between the two indices. The accuracy of recall of colonoscopy is in keeping with results from studies that assessed the validity and reliability of self-reported colorectal cancer screening, and found that colonoscopy is more accurately recalled than other colorectal cancer tests (5, 13, 14). This may be due to the lengthy and sometimes uncomfortable preparation regime required prior to colonoscopy.

Test-retest reliability was good for diagnosis date and for reported date of colonoscopy, but the validity for colonoscopy was not quite as strong: there was exact agreement between participants and GP data for only about half of the dates (although validity with the same month and year remained moderate to good). These results raise concerns about the quality of self-reported data on the process of diagnosis. Of course, the research question being addressed by a study would determine whether or not it is important that patients recall symptoms, test and

dates with great accuracy. For the purpose of examination of delay in diagnosis of colorectal cancer, our findings suggest that researchers should use self-reported data with caution.

A limitation of our study is that we used data collected by GP questionnaires as the “gold standard” comparison for validity analyses. We were unable to access GP records ourselves due to cost and time restrictions, as our sample resided across the large state of Queensland.

Medical records and physician notes are not always accurate or complete (5, 14, 16), and have been described as an unacceptable criterion against which to assess self-report (7). By asking GPs to complete a questionnaire based on their records we likely attributed an additional source of error/bias to the data. However, in practice it is often difficult and expensive to collect more objective sources of comparison data. For example, comparing reported blood tests against pathology reports would be costly and impractical, in the Australian context, as there are numerous pathology service providers involved in diagnostic testing. Further, not all diagnostic tests result in reports being generated, for example digital rectal examination and abdominal palpation, therefore GP records will sometimes be the only comparison available.

Another limitation was the collection of symptom data through spontaneous reporting. Funch (26) found that responses to open-ended questions underestimated the number of symptoms that colorectal cancer patients reported. Therefore participants in this study may have reported slightly different symptoms at each interview, these symptoms being different simply because they represented a proportion of all the symptoms the participant had experienced. It is likely that the reliability of reported symptoms would have been higher if the interviewer had read out an extensive list of symptoms, and asked participants whether they had experienced each of the symptoms. However, this method may also lead to over-reporting of symptoms, particularly those not specific to colorectal cancer, which may also jeopardise the validity of this type of measure.

Further research is needed to establish the reliability and validity of self-reported key elements in the process of diagnosis of different cancers. Disease characteristics, treatments administered and demographic characteristics of patients can vary widely, making generalization of findings inadvisable across cancer types. Different administration methods, such as written questionnaires and face-to-face interviews, should also be assessed to determine whether mode of administration alters the test-retest reliability and validity for these items, particularly in regard to the sourcing of data from GP records.

Although imperfect, self-reported information is an important means of collecting data for epidemiologic studies. Our findings provide some evidence to help guide the interpretation of self-reported symptoms, tests and dates in the colorectal cancer diagnostic pathway.

Acknowledgements

The authors wish to acknowledge the general practitioners who kindly completed the questionnaires for this study. This project was funded by The Cancer Council Queensland

References

1. Ferlay J, Bray F, Pisani P, Parkin D. GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide. Lyon: IARC Press, 2004.
2. Jemal A, Murray T, Ward E, et al. Cancer statistics, 2005. CA - A Cancer Journal for Clinicians 2005;55:10-30.
3. Redaelli A, Crano C, Okano G, PR R. Screening, prevention and socioeconomic costs associated with the treatment of colorectal cancer. Pharmacoeconomics 2003;21:1213-1238.
4. Shelton B. Introduction to colorectal cancer. Seminars in Oncology Nursing 2002;2:2-12.

5. Bradbury B, Brooks D, Brawarsky P, Mucci L. Test-retest reliability of colorectal testing questions on the Massachusetts Behavioral Risk Factor Surveillance System (BRFSS). *Preventive Medicine* 2005;41:303-311.
6. Caplan L, McQueen D, Qualters J, Leff M, Garrett C, Calonge N. Validity of women's self-reports of cancer screening test utilization in a managed care population. *Cancer Epidemiology, Biomarkers and Prevention* 2003;12:1182-1187.
7. Newell S, Girgis A, Sanson-Fisher R, Savolainen N. The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population: a critical review. *American Journal of Preventive Medicine* 1999;17:211-229.
8. Whitman S, Lacey L, Ansell D, Chen E, Dell J, Phillips C. Do chart reviews and interviews provide the same information about breast and cervical cancer screening. *International Journal of Epidemiology* 1993;22:393-397.
9. Champion V, Menon U, McQuillen D, Scott C. Validity of self-reported mammography in low-income African-American women. *American Journal of Preventive Medicine* 1998;14:111-117.
10. Coughlin S. Recall bias in epidemiologic studies. *Journal of Clinical Epidemiology* 1990;43:87-91.
11. Hancock L, Sanson-Fisher R, Kentish L. Cervical cancer screening in rural NSW: Health Insurance Commission data compared to self-report. *Australian and New Zealand Journal of Public Health* 1998;22:307-312.
12. Malats N, Belloc J, Gallen M, Porta M. Disagreement between hospital medical records and a structured patient interview on the type and date of the first symptom in cancers of the digestive tract. *Revue D Epidemiologie et de Sante Publique* 1995;43:533 - 540.
13. Baier M, Calonge N, Cutter G, et al. Validity of self-reported colorectal cancer screening behavior. *Cancer Epidemiology, Biomarkers and Prevention* 2000;9:229-232.

14. Hall H, Van Den Eeden S, Tolsma D, et al. Testing for prostate and colorectal cancer: comparison of self-report and medical record audit. *Preventive Medicine* 2004;39:27-35.
15. Hiatt R, Perez-Stable E, Quesenberry C, Sabogal F, Otero-Sabogal R, McPhee S. Agreement between self-reported early cancer detection practices and medical audits among Hispanic and non-Hispanic white health plan members in northern California. *Preventive Medicine* 1995;24:278-285.
16. Montano D, Phillips W. Cancer screening by primary care physicians: a comparison of rates obtained from physician self-report, patient survey, and chart audit. *American Journal of Public Health* 1995;85:795-800.
17. Lewis R. Reliability and validity: meaning and measurement. Annual Meeting of the Society for Academic Emergency Medicine. Boston, Massachusetts, 1999.
18. Fleiss J. *Statistical Methods for Rates and Proportions*. New York: Wiley, 1981.
19. Hripcsak G, Heitjan D. Measuring agreement in medical informatics reliability studies. *Journal of Biomedical Informatics* 2002;35:99-110.
20. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37-46.
21. Ebell M. *An Introduction to Information Mastery*. Lecture Notes, Michigan State University, 1999.
22. SAS Institute Inc. *SAS 9.1.3 Help and Documentation*. Cary, North Carolina: SAS Institute Inc, 2004.
23. Cicchetti D, Allison T. A new procedure for assessing reliability of scoring EEG sleep recordings. *American Journal of EEG Technology* 1971;11:101-109.
24. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;i:307-310.
25. Tukey J. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley, 1977.
26. Funch D. Predictors and consequences of symptom reporting behaviors in colorectal cancer patients. *Medical Care* 1988;26:1000-1008.

Legend

Tables

Table 1. Criteria used for interpreting indices of reliability and validity in this study

Table 2. Sociodemographic and physical characteristics of sample

Table 3. Reliability of symptoms and diagnostic tests reported by colorectal cancer patients:

Interview 1 vs Interview 2

Table 4. Validity of symptoms and diagnostic tests reported by colorectal cancer patients:

Interview 1 vs GP data

Figures

Figure 1. Bland-Altman plot of date of first GP consultation: Interview 1 vs Interview 2

Figure 2. Bland-Altman plot of date of first GP consultation: Interview 1 vs GP data

Table 1: Criteria used for interpreting indices of reliability and validity in this study

Reliability/Validity	Positive specific agreement	Kappa coefficient
High	> 80%	> 0.75
Moderate	50 – 80%	0.40 – 0.75
Low	< 50%	< 0.40

TABLE 2: Sociodemographic and disease-related characteristics of sample (n = 116)

Sample characteristics	n	(%)
Gender		
Male	70	(60.3)
Female	46	(39.7)
Age group		
< 50 years	7	(6.1)
50 – 59 years	15	(12.9)
60 – 69 years	47	(40.5)
≥ 70 years	47	(40.5)
Marital status		
Married or de facto	95	(81.9)
Widowed	13	(11.2)
Divorced or separated	6	(5.2)
Never married	2	(1.7)
Highest educational attainment		
< 12 years (did not complete high school)	72	(62.1)
12 years (completed high school)	14	(12.1)
Technical college	19	(16.4)
University	11	(9.4)
Health insurance status		
Private health insurance	59	(50.9)
No health insurance	57	(49.1)
Site of disease^a		
Colon	58	(67.4)
Rectum	28	(32.6)
Stage of disease^b		
Duke's A	33	(30.8)
Duke's B	39	(36.5)
Duke's C	26	(24.3)
Duke's D	9	(8.4)
Treatment type		
Surgery only	67	(57.7)
Surgery and adjuvant therapy	43	(37.1)
No treatment at this stage	6	(5.2)
Creation of stoma		
None	97	(83.6)
Temporary stoma	15	(12.9)
Permanent stoma	4	(3.5)

a missing data (n=30)

b missing data (n=9)

TABLE 3: Reliability of symptoms and diagnostic tests* reported by colorectal cancer patients: Interview 1 vs Interview 2

	Interview 1	Yes	No	Yes	No	Positive specific		Kappa coefficient	
	Interview 2	Yes	Yes	No	No	agreement (95% CI)		(95% CI)	
Symptoms (n = 116)									
Blood in stools		35	3	5	73	90	(83, 97)	0.85	(0.74, 0.95)
Pain in lower abdomen		20	6	5	85	78	(66, 91)	0.72	(0.57, 0.88)
Change in bowel habit		25	7	5	79	81	(70, 91)	0.74	(0.60, 0.88)
Lack of energy/tiredness		6	5	5	100	55	(29, 80)	0.50	(0.23, 0.77)
Diagnostic tests (n = 116)									
Digital rectal examination		13	5	12	86	60	(43, 78)	0.52	(0.32, 0.72)
Faecal occult blood test		10	2	5	99	74	(56, 93)	0.71	(0.50, 0.91)
Colonoscopy		95	4	8	9	94	(91, 97)	0.54	(0.31, 0.77)
Blood test		22	13	6	75	70	(57, 83)	0.59	(0.42, 0.75)
CT scan		6	4	5	101	57	(32, 82)	0.53	(0.26, 0.80)
Ultrasound		8	5	1	102	73	(52, 94)	0.70	(0.47, 0.92)
X-ray		9	1	4	102	78	(60, 97)	0.76	(0.56, 0.96)

* Note: Results not shown for the symptoms of pain elsewhere, increased flatulence, nausea or vomiting, unexplained weight loss, and loss of appetite, or for the flexible sigmoidoscopy diagnostic test due to insufficient data

TABLE 4: Validity of symptoms and diagnostic tests* reported by colorectal cancer patients: Interview 1 vs GP data

	Interview 1	Yes	No	Yes	No	Positive specific		Kappa coefficient	
	GP data	Yes	Yes	No	No	agreement (95% CI)		(95% CI)	
Symptoms (n = 92)									
Blood in stools		19	10	10	53	66	(51, 80)	0.50	(0.31, 0.69)
Pain in lower abdomen		11	5	10	66	59	(41, 78)	0.49	(0.28, 0.71)
Change in bowel habit		15	10	11	56	59	(43, 75)	0.43	(0.23, 0.64)
Lack of energy/tiredness		1	0	10	81	17	(0, 45)	0.15	(-0.11, 0.41)
Diagnostic tests (n = 95)									
Digital rectal examination		13	12	7	63	58	(41, 75)	0.45	(0.24, 0.66)
Faecal occult blood test		9	6	4	76	64	(44, 85)	0.58	(0.35, 0.81)
Colonoscopy		75	3	11	6	91	(87, 96)	0.39	(0.13, 0.64)
Blood test		13	8	11	63	58	(41, 75)	0.45	(0.24, 0.66)
CT scan		1	2	8	84	17	(0, 45)	0.12	(-0.16, 0.41)
Ultrasound		1	2	6	86	20	(0, 53)	0.16	(-0.17, 0.50)
X-ray		2	0	8	85	33	(0, 68)	0.31	(-0.02, 0.64)

* Note: Results not shown for the symptoms of pain elsewhere, increased flatulence, nausea or vomiting, unexplained weight loss and loss of appetite, or the flexible sigmoidoscopy diagnostic test due to insufficient data.

FIGURE 1: Bland-Altman plot of date of first GP consultation: Interview 1 vs Interview 2

FIGURE 2: Bland-Altman plot of date of first GP consultation: Interview 1 vs GP data

Self-reported information on the diagnosis of colorectal cancer:
reliable, but not necessarily valid

Word count (abstract): 193

Word count (text): 3,540