Cancer Epidemiology 82 (2023) 102310

Contents lists available at ScienceDirect

Cancer Epidemiology

journal homepage: www.elsevier.com/locate/canep

Examining methodology to identify patterns of consulting in primary care for different groups of patients before a diagnosis of cancer: An exemplar applied to oesophagogastric cancer

Sarah Price, Bianca Wiering *, Luke T.A. Mounce, Willie Hamilton, Gary Abel

Medical School, College of Medicine and Health, University of Exeter, St Luke's Campus, Heavitree Road, Exeter EX1 2LU, UK

ARTICLE INFO

Keywords:
Early diagnosis of cancer
Methodological study
Maximum likelihood estimates

ABSTRACT

Background: Current methods for estimating the timeliness of cancer diagnosis are not robust because dates of key defining milestones, for example first presentation, are uncertain. This is exacerbated when patients have other conditions (multimorbidity), particularly those that share symptoms with cancer. Methods independent of this uncertainty are needed for accurate estimates of the timeliness of cancer diagnosis, and to understand how multimorbidity impacts the diagnostic process.

Methods: Participants were diagnosed with oesophagogastric cancer between 2010 and 2019. Controls were matched on year of birth, sex, general practice and multimorbidity burden calculated using the Cambridge Multimorbidity Score. Primary care data (Clinical Practice Research Datalink) was used to explore population-level consultation rates for up to two years before diagnosis across different multimorbidity burdens. Five approaches were compared on the timing of the consultation frequency increase, the inflection point for different multimorbidity burdens, different aggregated time-periods and sample sizes.

Results: We included 15,410 participants, of which 13,328 (86.5 %) had a measurable multimorbidity burden. Our new maximum likelihood estimation method found evidence that the inflection point in consultation frequency varied with multimorbidity burden, from 154 days (95 %CI 131.8–176.2) before diagnosis for patients with no multimorbidity, to 126 days (108.5–143.5) for patients with the greatest multimorbidity burden. Inflection points identified using alternative methods were closer to diagnosis for up to three burden groups. Sample size reduction and changing the aggregation period resulted in inflection points closer to diagnosis, with the smallest change for the maximum likelihood method.

Discussion: Existing methods to identify changes in consultation rates can introduce substantial bias which depends on sample size and aggregation period. The direct maximum likelihood method was less prone to this bias than other methods and offers a robust, population-level alternative for estimating the timeliness of cancer diagnosis.

1. Introduction

Early cancer diagnosis remains a focus of UK policy and research, with time to diagnosis a key outcome [1,2]. Standardised definitions of time points and intervals describe patients' pre-diagnostic pathways, some of which are objective (e.g. diagnosis date) while others are subjective, such as date of first presentation [3]. Symptoms of possible cancer commonly have other causes [4], particularly in people with two or more chronic conditions (i.e. multimorbidity) [5]. Multimorbidity also more than doubles the primary-care consultation rate [6,7], increasing the chance that possible cancer symptoms are recorded.

Modelling population-level consultation rates, which rise before a cancer diagnosis, may offer a robust alternative to patient-level metrics. Existing methods identify statistically significant deviations in consultation rate, either between groups or from historical trends [8]. Statistical significance depends on effect and sample size; therefore, the time of consultation-rate change may vary with group size or underlying consultation rate. Furthermore, these methods cannot quantify the uncertainty around the timing of the deviation.

We urgently need workable and accurate metrics of the timeliness of cancer diagnosis that are independent of such biases and robust in patients with multimorbidity, who represent over three-quarters of

* Correspondence to: Smeall Building, University of Exeter, St Luke's Campus, Heavitree Road, Exeter EX1 2LU, UK.
E-mail address: b.wiering@exeter.ac.uk (B. Wiering).

https://doi.org/10.1016/j.canep.2022.102310
Received 31 August 2022; Received in revised form 25 November 2022; Accepted 30 November 2022
Available online 9 December 2022

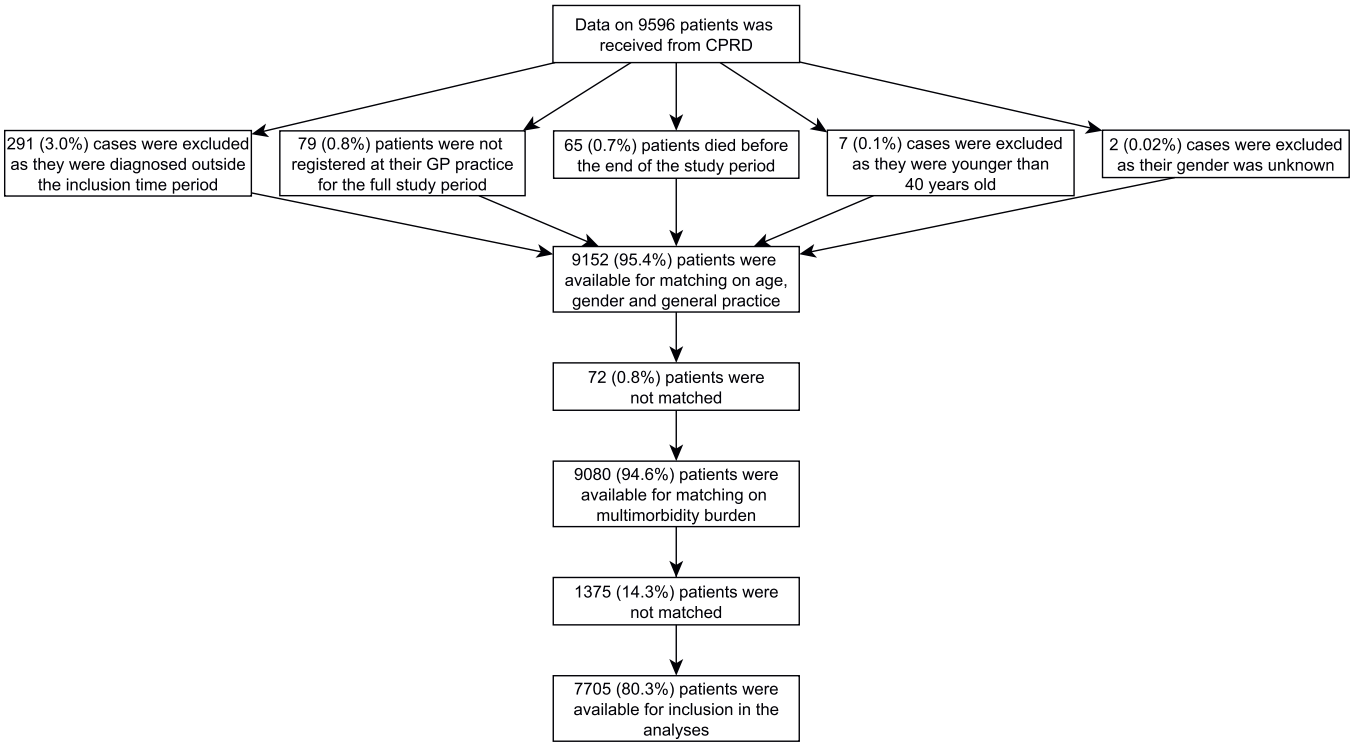


Fig. 1. Flowchart of case selection and matching process.

patients aged ≥ 75 years – the peak age of cancer incidence [7,9]. We explore using population-level consultation rates in primary care before cancer diagnosis as a measure of diagnostic timeliness across groups of patients with different multimorbidity burden. For two new methods and three approaches used previously [10–12] we:

1. Identify the time before cancer diagnosis that primary-care consulting frequency increases above the norm (i.e. the inflection point).
2. Compare the inflection point between patients with different multimorbidity burden.
3. Investigate potential biases introduced by varying the period over which consultations are aggregated (28 days vs. 7, 14, and 21 days) and the sample size (100 % vs 50 %, 20 %, 10 %, and 5 %).

We illustrate the methods using oesophagastric cancer, which presents with a broad range of non-specific symptoms [13] that commonly feature in chronic conditions [14].

2. Materials and methods

2.1. Study design and data source

This matched, retrospective cohort study analysed primary-care consultation rates of cases and controls for 2 years before the case's oesophagogastric cancer diagnosis. Cases and matched controls were continuously registered at a Clinical Practice Research Datalink (CPRD) practice with up-to-standard data for the 2 years before diagnosis.

The data source was CPRD GOLD, a routinely collected UK primary-care database of medical records, with high data quality and validity, covering 8 % of the UK [15]. CPRD GOLD, which contains patient demographic data and clinical information, is frequently used for cancer diagnostic studies [16–22].

2.2. Participants and matching

We requested all CPRD GOLD participants (n = 9596) aged ≥ 40 years who had diagnostic Read codes for oesophagastric cancer (equivalent to International Classification of Disease codes C15 or C16) between 1 January 2010 and 31 December 2019. Participant characteristics were similar to those of patients diagnosed with oesophagogastric cancer in England between 2018 and 2020 [23]. Cases were first matched with a maximum of up to 20 controls (n = 189,673), who did not have oesophagogastric cancer codes and were randomly selected from CPRD GOLD, on year of birth, sex, and general practice in order to generate a patient-level multimorbidity burden variable. Patient-level multimorbidity burden was defined as the overall impact of a range of conditions present before the cancer diagnosis, and estimated using the Cambridge Multimorbidity Score's general outcome weighting [24]. Participants not meeting Cambridge Multimorbidity Score criteria were assumed to have no multimorbidity burden. Four multimorbidity burden groups (no, low, medium or high) were derived, the last three from Cambridge Multimorbidity Score tertiles.

Final case:control matching was 1:1 on year of birth, sex, general practice and multimorbidity burden group. Where multiple controls matched to a case, one was selected at random without replacement (Fig. 1).

2.3. Outcome and covariates

Patient-level consultation rate was the number of consultation-days per aggregation period. A 28-day aggregation period approximated month before diagnosis (see section 2.5.6 Extra analyses for aggregation period variation). A consultation was defined as any GP visit coded in CPRD as a face-to-face or telephone consultation (Table A1, Appendix 1). A consultation-day was one when the participant had at least one such CPRD consultation code.

2.4. Dataset construction

We created a panel dataset with unique identifiers for individual *participants* and *general practices*, and variables identifying *sex*, *age*, *case–control status*, *matched case–control pairs*, and *morbidity burden*

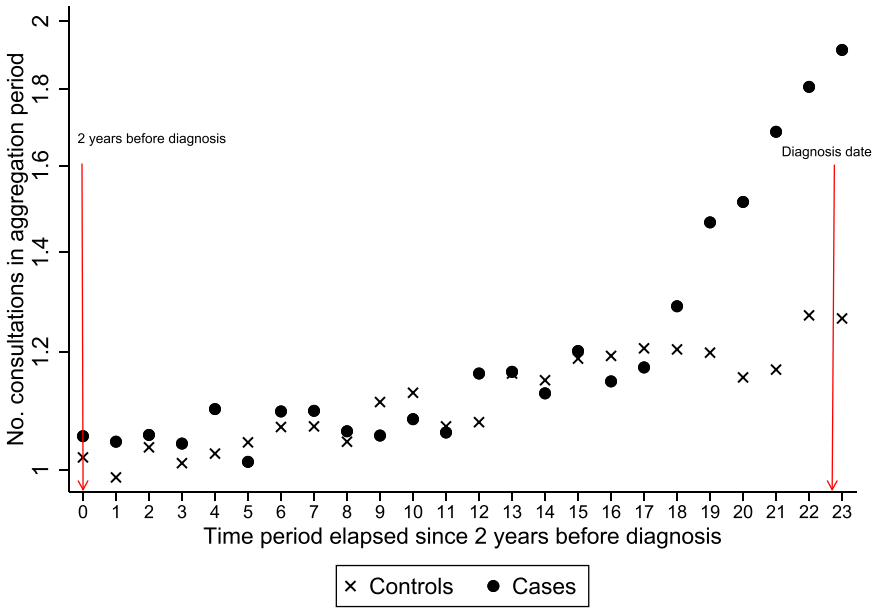


Fig. 2. Simulated dataset of 1000 cases and 1000 matched controls, illustrating the naming of time-period variables relative to diagnosis date.

group. A dummy variable represented the discrete *time-period* counting up to diagnosis, which took a value of 0 at 2 years before diagnosis, and was incremented by 1 for each aggregation period until the diagnosis date. For a 28-day aggregation period, the *time-period* variable ranged from 0 to 23, giving 24 rows per patient. A *count* variable quantified the patient-level number of consultation-days in each time-period (see Fig. 2 for simulated data), which formed the basis for the dependent variables modelled. Where controls were included in the models, the same time-period was used to calculate the controls' consultation frequency as for their matched case. Other covariates were created for the methods described below.

2.5. Statistical analyses

2.5.1. Direct maximum likelihood method

We derive a method for estimating the inflection point (i.e. the time before diagnosis when the consultation rate increases above the baseline trend) with 95 % confidence intervals, using mixed-effects, negative binomial time-series regression analyses similar to methods employed by our team in recent studies [25,26]. This method required 24 unique *inflection-point* variables, to test whether each aggregation period included the inflection point. These discrete time variables were set to 0 for controls. For cases, they were derived from the *time-period* variable by subtracting an amount equal to the number of the period being tested, and had a minimum value of 0. For example, the *inflection-point* variable for testing time-period 2 took values of 0 for time-periods 0, 1, and 2, after which it was incremented by 1 to a maximum of 21 in time-period 23 (i.e. the period of diagnosis). (See Mendeley Data link 10.17632/3mj526hgzx.3 for Stata code to create these variables and an example of the dataset construction [27]).

Analyses were stratified by multimorbidity burden group, with 24 mixed-effects, negative binomial time-series regression models run per group. Models employed a random intercept for matched pair, and modelled the log of the expected *number of consultation-days* in 28-day aggregation periods. The covariates were *time-period*, *case–control* status, and a single *inflection-point* variable representing the time-period being tested as containing the inflection point. The coefficient for time-period quantified the secular trend in consultation rate. Using an interaction term between case-control status and time-period allowed us to subtract the secular trend in consultation rate from the difference between cases and controls. We ran 24 models, each containing an interaction term between case–control status and the time-period being tested as containing the inflection-point. This interaction captured any differential change in the rate of consultations between cases and controls around the inflection point. The model with the highest maximum likelihood estimate was assumed to provide the best fit to the data. The time-period used in that model was chosen as the period most likely to contain the inflection point (see Mendeley Data link 10.17632/3mj526hgzx.3 [27] for the Stata syntax for the models).

The model selection process was bootstrapped (n = 50, with sampling clustered by matched pair) to obtain 95 % confidence intervals on the period containing the inflection point. We also quantified the difference in inflection point (95 %CI) between multimorbidity burden groups (reference group, no multimorbidity burden).

2.5.2. Comparison method A: case-control comparison over time, controlled for baseline rate differences between cases and controls

This method identified the earliest point before diagnosis at which the consultation rate is significantly greater for cases than controls, allowing for secular trends in consultation rate [28]. Mixed-effects, time-series negative binomial regression models, stratified by multimorbidity burden, employed a random intercept for matched pair, and used the clustered sandwich estimator to relax the requirement for independence of observations within practices. The outcome was number of consultations, and explanatory variables were *case-control status* and 24 interaction terms between *case–control status* and *time-period*. The reference time-period was the furthest from diagnosis, i.e. period 0. The interaction terms report how much the effect of having undiagnosed cancer on consultation rate differs by time, controlling for differences in the baseline incidence rate between cases and controls [29]. We selected the inflection point as the earliest time-period before diagnosis that the interaction term became consistently (two or more consecutive time-periods) significant ($p < 0.0001$).

2.5.3. Comparison method B: month-by-month case control comparison

This method compared the number of consultations between cases and controls in each discrete aggregation period before diagnosis [8]. For a 28-day aggregation period, 96 separate models were run: one for each time-period and multimorbidity burden level. We used three-level mixed-effects, negative binomial regression models, employing random intercepts for matched pair nested within general practice. The outcome was number of consultations, and the single covariate was *case–control*

Table 1
Participant (50:50 cases and controls) characteristics, by burden group.

Burden group[a]	N (% male)	Mean age (SD)
None	2082 (70.4)	63.0 (10.1)
Low	3786 (67.4)	69.3 (10.4)
Medium	3856 (66.2)	72.3 (10.0)
High	5686 (64.5)	76.6 (8.89)

[a] The Low, Medium, and High burden groups were derived using Cambridge Multimorbidity Score tertiles [32] (see Section 2.2).

status. We selected the inflection point as the earliest time-period that the mean consultation rate was consistently (two or more consecutive time-periods) different between cases and controls ($p < 0.0001$).

2.5.4. Comparison methods C and D: comparison against baseline and comparison against previous month

These methods analysed data from cases only [8]. We modelled the log of the expected *number of consultation-days* in the 28-day aggregation period as a function of discrete time before diagnosis. Models, stratified by multimorbidity burden, were adjusted for age and sex. Two methods identified the inflection point:

Method C identified the earliest *time-period* where the number of consultations was greater than the baseline value at 2 years before diagnosis (at $p < 0.0001$).

Method D identified the first time-period when the number of consultations exceeded that in the previous one, by performing pairwise comparisons of the predicted marginal consultation rates across all 24 time-periods. Bonferroni correction reduced the risk of type 1 errors.

2.5.5. Converting time-period to days before diagnosis

The inflection point is reported as days before diagnosis by:

1. Identifying the time-period in terms of time *before* diagnosis
2. Multiplying by aggregation period duration.
3. Adding half an aggregation period to reflect the middle of the period.

For example, using a 28-day aggregation period and 24 time-periods, of which time-period 19 is identified as most likely to contain the inflection point:

Time-period *before* diagnosis = 24–19 = 5.

Days before diagnosis = $5 \times 28 + 14 = 154$.

For the maximum likelihood method, the 95 % confidence intervals were also converted to days before diagnosis using these steps.

2.5.6. Extra analyses

To explore for potential bias, we repeated all methods using 5 %, 10 %, 20 % or 50 % of the original sample size. Random subsamples of matched pairs were drawn from the main dataset without replacement.

We examined the effect of varying the aggregation period to 7, 14, or 21 days. Fresh panel datasets with new *time-period* and *inflection-point* variables were created.

2.6. Ethics approval

Ethics approval was granted for all observational research using anonymised CPRD data by the National Research Ethics Service (NRES) [15]. The Independent Scientific Advisory Committee approved the study protocol (20_124) on 23 June 2020.

3. Results

The final cohort numbered 15,410 participants (66.4 % male) (Table 1). Mean age increased with rising multimorbidity burden, and the percentage who were male decreased.

At 24 months before diagnosis, consultation-days per 28-day aggregation period were similar for cases and controls within burden group (Fig. 3). Consultation rate increased with multimorbidity burden (units are consultation-days per person per 28-day aggregation period). It was 0.21 (range: 0–8; SD: 0.61) for cases and 0.21 (range 0–10, SD 0.69) for controls at burden level 0, rising to 1.02 (range 0–19; SD 1.38) for cases and 0.98 (range: 0–18; SD 1.32) for controls at burden level 3.

3.1. Inflection points across the different methods, 28-day aggregation period

The direct maximum likelihood method estimated the inflection points to be 154 (95 %CI 131.9–176.2), 126 (103.0–149.0), 126 (115.1–136.9) and 126 (108.5–143.5) days before diagnosis,

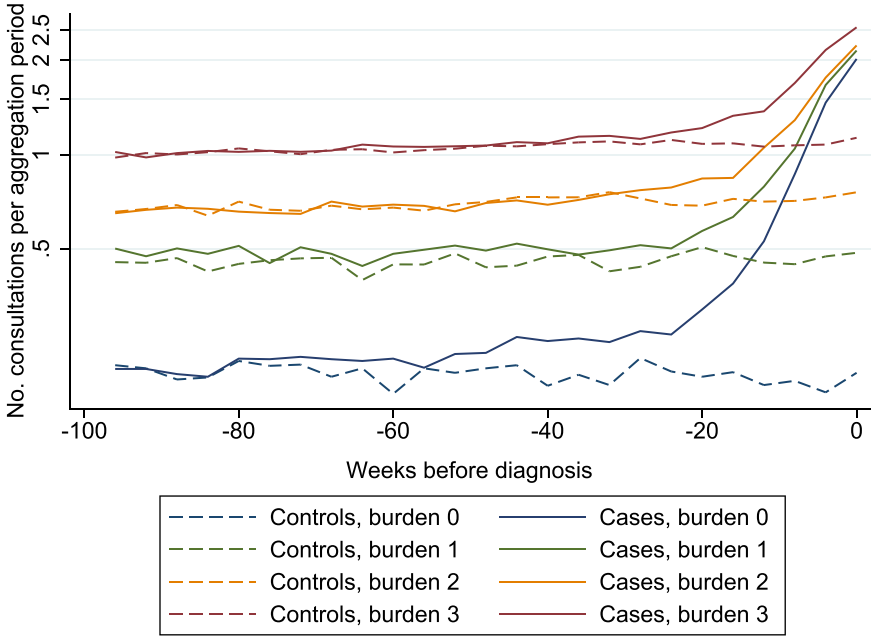
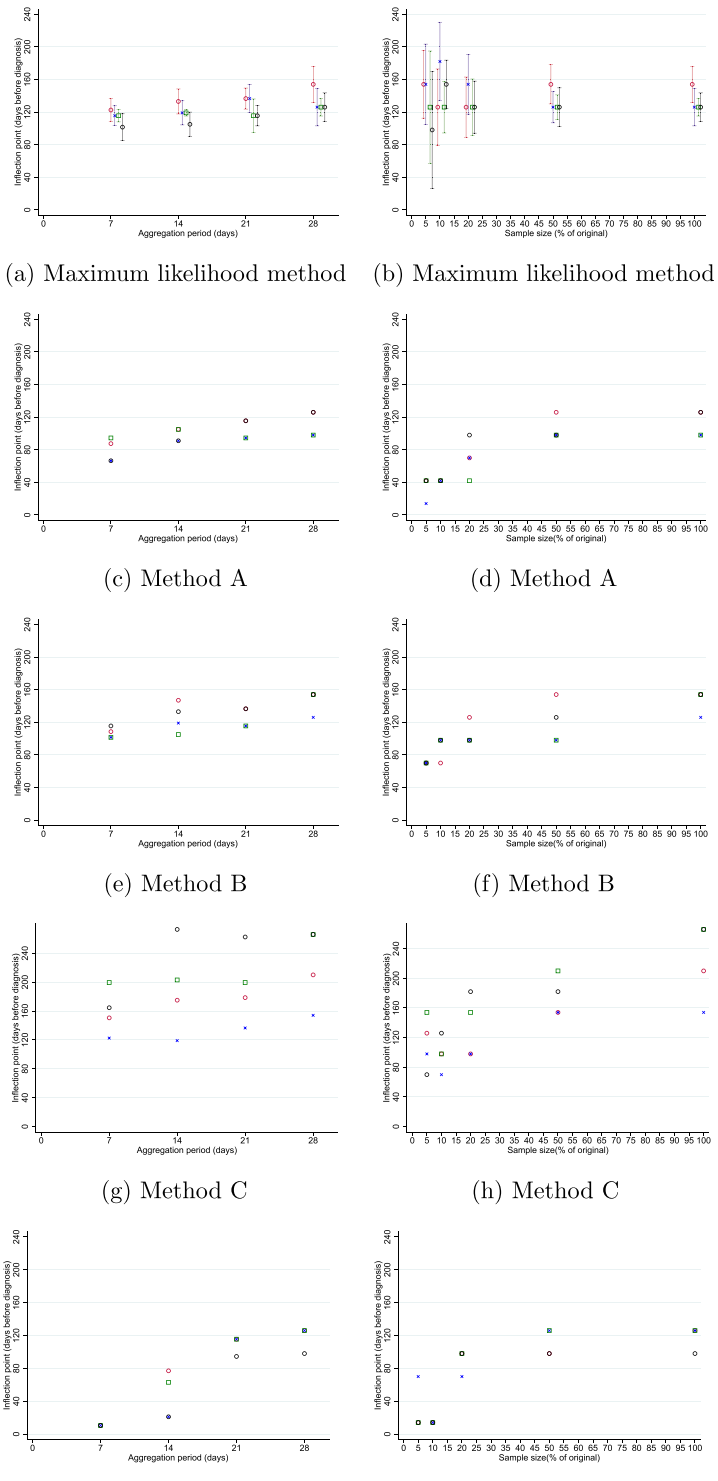


Fig. 3. Raw data plot of consultation-days per 28-day aggregation period for cases and controls by burden group in the 2 years before the case is diagnosed with oesophagogastric cancer.



(a) Maximum likelihood method (b) Maximum likelihood method

(c) Method A (d) Method A

(e) Method B (f) Method B

(g) Method C (h) Method C

(i) Method D (j) Method D

Fig. 4. Estimation of inflection point (days before diagnosis), by burden group (group 0: red circles; group 1: blue crosses; group 2: green squares; group 3: black circles), for aggregation periods (left panels) of 7, 14, 21 and 28 days, and for 5, 10, 20, 50 % and 100 % sample size (right panels). **a, b** Maximum likelihood method; **c, d** Method A: Case–control comparison, controlled for baseline rate differences between cases and controls; **e, f** Method B: Case–control comparison; **g, h** Method C: Comparison with baseline; **i, j** Method D: Comparison with previous month.

respectively, for burden levels 0, 1, 2, and 3 (Fig. 4a, Table 2). There was moderate evidence that the inflection point for burden group 0 occurred 28 days (95 %CI 4.3–51.7, p = 0.02) earlier than for burden group 2.

Inflection points estimated by Method A were 28 days closer to diagnosis than those estimated by the direct maximum likelihood method for all burden levels except level 3 (Table 2, Fig. 4a, c).

Mean consultation rate differed between cases and controls (Method B) at a time consistent with the inflection point identified by the direct maximum likelihood method for burden groups 0 and 1. For burden groups 2 and 3, it occurred earlier (154 days before diagnosis – outside the 95 %CI estimated by the maximum likelihood method) (Table 2, Fig. 4a, e).

For all burden levels, the consultation rate by cases exceeded that in the 24th month before diagnosis (Method C) at a time consistently earlier than reported by the direct maximum likelihood method (Table 2, Fig. 4a, g).

The predicted marginal consultation rate for cases changed (Method D) at a time consistent with the maximum likelihood method's estimate

Table 2
Inflection point estimations for all methods for smaller time aggregations and a 100 % sample.

Method	Multi-morbidity Burden level	Inflection point estimate (days before diagnosis) for aggregation periods of:			
		28 Days	21 Days	14 Days	7 Days
Direct maximum likelihood	0	154	137	133	123
	1	126	137	119	116
	2	126	116	119	116
	3	126	116	105	102
Method A	0	126	116	105	88
	1	98	95	91	67
	2	98	95	105	95
	3	126	116	91	67
Method B	0	154	137	147	109
	1	126	116	119	102
	2	154	116	105	102
	3	154	137	133	116
Method C	0	210	179	175	151
	1	154	137	119	123
	2	266	200	203	200
	3	266	263	273	165
Method D	0	126	116	77	11
	1	126	116	21	11
	2	126	116	63	11
	3	98	95	21	11

Table 3
Inflection point estimations for all methods for smaller sample sizes and a 28-day aggregation period.

Methods	Multi-morbidity Burden level	Percentage of original sample size included:			
		5 %	10 %	20 %	50 %
Direct maximum likelihood	0	154	126	126	154
	1	154	182	154	126
	2	126	126	126	126
	3	98	154	126	126
Method A	0	42	42	70	126
	1	14	42	70	98
	2	42	42	42	98
	3	42	42	98	98
Method B	0	70	70	126	154
	1	70	98	98	98
	2	70	98	98	98
	3	70	98	98	126
Method C	0	126	98	98	154
	1	98	70	98	154
	2	154	98	154	210
	3	70	126	182	182
Method D	0	14	14	98	98
	1	70	14	70	126
	2	14	14	98	126
	3	14	14	98	98

of inflection point for burden levels 1 and 2. For burden levels 0 and 3, this point occurred closer to diagnosis than estimated by the maximum likelihood method (Table 2, Fig. 4a, i).

3.2. Varying sample size

For the maximum likelihood method, reducing the sample size below 50 % widened the 95 % confidence intervals (Fig. 4b). However, there was no notable trend in the point estimates of inflection points with reducing sample size. The greatest variation for other burden levels occurred when the sample was reduced to 20 % or more of its original size. For example, a decrease to 98 days before diagnosis for burden level 3, and an increase to 154 days for burden level 1 for 5 % of the total sample.

For comparison method A, the inflection point estimate moved closer to the diagnosis date with sample sizes at or below 50 % of the original sample (Table 3, Fig. 4d).

Estimates of when mean consultation rate differed between cases and controls (Method B) and of when cases consulted more than in the 24th month before diagnosis (Method C) tended to move closer to the diagnosis date with reducing sample size (Table 3, Fig. 4f, h).

Method D was unable to estimate the time when the predicted marginal consultation rate was statistically significantly different from the month before when sample size was reduced to 10 % and 5 %.

3.3. Varying aggregation period

For the maximum likelihood method, the confidence interval width remained similar across reducing aggregation periods. For all methods, the inflection point estimation moved closer to the diagnosis date with shorter aggregation periods, notably at or less than 14 days (Table 2, Fig. 4a, c, e, g, i). The change in inflection point varied by method, being smallest for the maximum likelihood method (up to 24 days) and largest for comparison method D (115 days).

4. Discussion

4.1. Summary of findings

We have developed a method for robustly estimating when primary-care consultation rates rise above the norm before a cancer diagnosis. This is a population-based measure of diagnostic timeliness. We demonstrated that existing alternative methods to identify changes in consultation rates can introduce substantial bias that depends on sample size and aggregation period. The new method was less prone to this bias than other methods. We found moderate evidence that this inflection point varied with multimorbidity burden. It occurred at 154 days before diagnosis for people with no multimorbidity, and at 126 days before diagnosis for patients with the greatest multimorbidity burden. Our current research will further explore the clinical implications of differences in inflection point by multimorbidity burden.

4.2. Strengths and limitations

This methodology was developed using CPRD GOLD, a large database of anonymised patient records widely used for primary-care and epidemiology studies [15]. Established methods identified cases [30]. Primary-care consultations are coded in the CPRD, overcoming the problems of missing data inherent to analyses of time intervals based on symptom records [31]. Confounding was minimised by matching on year of birth, sex, general practice and multimorbidity burden. We estimated multimorbidity burden using The Cambridge Multimorbidity Score, which outperforms the Charlson comorbidity index when predicting primary care consultations, unplanned hospitalisation, and mortality [32,33]. The Cambridge Multimorbidity Score is based on Quality and Outcomes Framework diagnoses, which general practitioners are incentivised to record, reducing the chance of missing data.

This study did not validate the cancer diagnoses with linked Cancer Registry data. The concordance between the two data sources suggests a low rate of case misidentification, although the CPRD diagnosis date was a median of 12 days after the Registry date, which has potential to affect estimates of the inflection point [34,35]. However, the results found using the Maximum likelihood method were similar to a review suggesting that change in health care use was observable within 6 months before a diagnosis of oesophagogastric cancers [8].

4.3. Comparison of methods

In making any recommendation for a methodological change, it is important to consider what one's analytical method is intended to do. In this paper, the method is designed to address flaws in current methods for assessing cancer diagnostic *system* change. If an intervention aiming to expedite cancer diagnosis in symptomatic cancer is successful, it is reasonable to infer that the time between first presentation of cancer symptoms to healthcare and diagnosis will have reduced. At an individual level, this period is called the diagnostic interval, and has historically been the main measure for such assessments. However, identifying the first 'milestone' on this process – the first presentation to healthcare – is fraught with difficulty, especially in patients with additional morbidities. Furthermore, such multimorbid patients are now the norm. Our use of an increase in consultation frequency (when compared to controls) is a reasonable proxy for the first presentation to healthcare, albeit only on a population basis.

If this argument is accepted, then our new Maximum Likelihood method has several advantages over existing methods. As it uses maximum likelihood estimation to identify the inflection point, it is not dependent on direct statistically significant comparisons. It also allows: calculation of confidence intervals; testing of differences between groups (in this example, morbidity burden levels), and it accounts for secular trends in baseline consultation rate. Although it is possible to adapt the maximum likelihood method to be used on a case-only dataset [25], without controls background trends are only inferred from the pre-inflection point period. It remains an open research question as to the degree to which this influences findings. Besides the maximum likelihood method, method A also takes into account a secular trend by controlling for differences in the baseline incidence rate between cases and controls, but its results are likely to be affected by temporary increases in consultation frequency because background trends are not modelled linearly. Additionally, and potentially most importantly, the maximum likelihood method is not prone to biases introduced by shorter aggregation periods and smaller sample sizes. Such biases risk erroneous conclusions being drawn if sample sizes vary between groups.

However, the maximum likelihood method is complex to use, and would need amending to account for any clustering within general practices. It also uses considerable computer resources, and takes longer to perform (about a day depending on the complexity of the model).

Even so, the advantages seem to outweigh disadvantages, though this is less important when a large sample is available. In this paper, we have applied the method to consultation rates, although it could be applied to any countable occurrences, such as prescriptions or tests.

5. Conclusions

The new method uses mixed-effects, negative binomial time series regression analyses to estimate the inflection point with 95 % confidence intervals, enabling a consistent estimation of the timeliness of cancer diagnosis at a population level. The method is less prone to bias than existing methods. Due to its complexity, other methods, such as Method A, may be considered for large studies with 28-days aggregation periods.

Funding

This research was funded by the National Institute for Health and Care Research (NIHR) Policy Research Programme, Policy Research Unit in Cancer Awareness, Screening and Early Diagnosis (grant number PRU-1217-21601). This research is also linked to the CanTest Collaborative, which is funded by CRUK (grant reference number: C8640/A23385), of which Willie Hamilton is Co-Director and Gary Abel is senior faculty. The funders did not have any role in the collection, analysis and interpretation of data; in the writing of the manuscript; or in the decision to submit the manuscript for publication. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Table A1
Classification of face-to-face and telephone consultations, based on the lookup for the "constype" variable in the CPRD Consultation file.

1	Clinic	FACE TO FACE/TELEPHONE
2	Night visit, Deputising service	FACE TO FACE/TELEPHONE
3	Follow-up/routine visit	FACE TO FACE/TELEPHONE
4	Night visit, Local rota	FACE TO FACE/TELEPHONE
5	Mail from patient	NOT ELIGIBLE
6	Night visit, practice	FACE TO FACE/TELEPHONE
7	Out of hours, Practice	FACE TO FACE/TELEPHONE
8	Out of hours, Non Practice	FACE TO FACE/TELEPHONE
9	Surgery consultation	FACE TO FACE/TELEPHONE
10	Telephone call from a patient	FACE TO FACE/TELEPHONE
11	Acute visit	FACE TO FACE/TELEPHONE
12	Discharge details	NOT ELIGIBLE
13	Letter from Outpatients	NOT ELIGIBLE
14	Repeat Issue	NOT ELIGIBLE
15	Other	NOT ELIGIBLE
16	Results recording	NOT ELIGIBLE
17	Mail to patient	NOT ELIGIBLE
18	Emergency Consultation	FACE TO FACE/TELEPHONE
19	Administration	NOT ELIGIBLE
20	Casualty Attendance	NOT ELIGIBLE
21	Telephone call to a patient	FACE TO FACE/TELEPHONE
22	Third Party Consultation	NOT ELIGIBLE
23	Hospital Admission	NOT ELIGIBLE
24	Children's Home Visit	FACE TO FACE/TELEPHONE
25	Day Case Report	NOT ELIGIBLE
26	GOS18 Report	NOT ELIGIBLE
27	Home Visit	FACE TO FACE/TELEPHONE
28	Hotel Visit	FACE TO FACE/TELEPHONE
29	NHS Direct Report	NOT ELIGIBLE
30	Nursing Home Visit	FACE TO FACE/TELEPHONE
31	Residential Home Visit	FACE TO FACE/TELEPHONE
32	Twilight Visit	FACE TO FACE/TELEPHONE
33	Triage	FACE TO FACE/TELEPHONE
34	Walk-in Centre	NOT ELIGIBLE
35	Co-op Telephone advice	NOT ELIGIBLE
36	Co-op Surgery Consultation	NOT ELIGIBLE
37	Co-op Home Visit	NOT ELIGIBLE
38	Minor Injury Service	NOT ELIGIBLE
39	Medicine Management	NOT ELIGIBLE
40	Community Clinic	NOT ELIGIBLE
41	Community Nursing Note	NOT ELIGIBLE
42	Community Nursing Report	NOT ELIGIBLE
43	Data Transferred from other system	NOT ELIGIBLE
44	Health Authority Entry	NOT ELIGIBLE
45	Health Visitor Note	NOT ELIGIBLE
46	Health Visitor Report	NOT ELIGIBLE
47	Hospital Inpatient Report	NOT ELIGIBLE
48	Initial Post Discharge Review	FACE TO FACE/TELEPHONE
49	Laboratory Request	FACE TO FACE/TELEPHONE
50	Night Visit	FACE TO FACE/TELEPHONE
51	Radiology Request	FACE TO FACE/TELEPHONE
52	Radiology Result	NOT ELIGIBLE
53	Referral Letter	FACE TO FACE/TELEPHONE
54	Social Services Report	NOT ELIGIBLE
55	Telephone Consultation	FACE TO FACE/TELEPHONE
56	Template Entry	FACE TO FACE/TELEPHONE
57	GP to GP communication transaction	NOT ELIGIBLE
58	Non-consultation medication data	NOT ELIGIBLE
59	Non-consultation data	NOT ELIGIBLE
60	ePharmacy message	NOT ELIGIBLE

CRediT authorship contribution statement

The study was conceived by GA. BW and SP curated the data. BW and SP analysed data and produced the first draft of the paper with supervision of GA, LM and WH. All authors interpreted findings and identified issues for discussion. The draft was reviewed and modified with input from all authors over a number of versions. All authors saw and approved the final version. SP and BW are guarantors.

Acknowledgements

We thank Professor Jose Valderas for his work classifying which consultations qualify as face-to-face or telephone consultations.

Declarations of interest

Willie Hamilton is Co-Director and Gary Abel is senior faculty of CanTest. The authors declare no further conflict of interest.

Appendix A

Table A1.

References

[1] Independent Cancer Taskforce, Achieving World-Class Cancer Outcomes: Taking the Strategy Forward, NHS England, London, 2016.
[2] Y. Romero, D. Trapani, S. Johnson, Z. Tittenbrun, L. Given, K. Hohman, L. Stevens, J.S. Torode, M. Boniol, A.M. Ilbawi, National cancer control plans: a global analysis, Lancet Oncol. 19 (10) (2018) e546–e555.
[3] D. Weller, P. Vedsted, G. Rubin, F.M. Walter, J. Emery, S. Scott, C. Campbell, R. S. Andersen, W. Hamilton, F. Olesen, P. Rose, S. Nafees, E. van Rijswijk, S. Hiom, C. Muth, M. Beyer, R.D. Neal, The Aarhus statement: improving design and reporting of studies on early cancer diagnosis, Br. J. Cancer 106 (7) (2012) 1262–1267.
[4] M. Biswas, A.E. Ades, W. Hamilton, Symptom lead times in lung and colorectal cancers: what are the benefits of symptom-based approaches to early diagnosis? Br. J. Cancer 112 (2) (2015) 271–277.
[5] C. Renzi, A. Kaushal, J. Emery, W. Hamilton, R.D. Neal, B. Rachet, G. Rubin, H. Singh, F.M. Walter, N.J. de Wit, Comorbid chronic diseases and cancer diagnosis: disease-specific effects and underlying mechanisms, Nat. Rev. Clin. Oncol. 16 (12) (2019) 746–761.
[6] C. Salisbury, L. Johnson, S. Purdy, J.M. Valderas, A.A. Montgomery, Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study, Br. J. Gen. Pract.: J. R. Coll. Gen. Pract. 61 (582) (2011) e12–e21.
[7] A. Cassell, D. Edwards, A. Harshfield, K. Rhodes, J. Brimicombe, R. Payne, S. Griffin, The epidemiology of multimorbidity in primary care: a retrospective cohort study, Br. J. Gen. Pract. 68 (669) (2018) e245–e251.
[8] B. White, C. Renzi, M. Rafiq, G.A. Abel, H. Jensen, G. Lyratzopoulos, Does changing healthcare use signal opportunities for earlier detection of cancer? A review of studies using information from electronic patient records, Cancer Epidemiol. 76 (2022), 102072.
[9] Cancer Research UK, Cancer Statistics for the UK, 2021. ⟨https://www.cancerresearchuk.org/health-professional/cancer-statistics-for-the-uk#heading-Zero⟩.
[10] P.L. Hansen, P. Hjertholm, P. Vedsted, Increased diagnostic activity in general practice during the year preceding colorectal cancer diagnosis, Int. J. Cancer 137 (3) (2015) 615–624.
[11] H. Jensen, P. Vedsted, H. Moller, Consultation frequency in general practice before cancer diagnosis in relation to the patient's usual consultation pattern: a population-based study, Cancer Epidemiol. 55 (2018) 142–148.
[12] L.M. Guldbrandt, H. Møller, E. Jakobsen, P. Vedsted, General practice consultations, diagnostic investigations, and prescriptions in the year preceding a lung cancer diagnosis, Cancer Med. 6 (1) (2017) 79–88.
[13] M.M. Koo, W. Hamilton, F.M. Walter, G.P. Rubin, G. Lyratzopoulos, Symptom signatures and diagnostic timeliness in cancer patients: a review of current evidence, Neoplasia (N. Y. N. Y.) 20 (2) (2018) 165–174.
[14] S. Stapley, T.J. Peters, R.D. Neal, P.W. Rose, F.M. Walter, W. Hamilton, The risk of oesophago-gastric cancer in symptomatic patients in primary care: a large case-control study using electronic records, Br. J. Cancer 108 (1) (2013) 25–31.
[15] E. Herrett, A.M. Gallagher, K. Bhaskaran, H. Forbes, R. Mathur, T. van Staa, L. Smeeth, Data Resource Profile: Clinical Practice Research Datalink (CPRD), Int. J. Epidemiol. 44 (3) (2015) 827–836.
[16] E.A. Shephard, S. Stapley, R.D. Neal, P. Rose, F.M. Walter, W.T. Hamilton, Clinical features of bladder cancer in primary care, Br. J. Gen. Pract. 62 (602) (2012) e598–e604.
[17] S. Stapley, T. Peters, R.D. Neal, P. Rose, F. Walter, W. Hamilton, The risk of oesophago-gastric cancer in symptomatic patients in primary care: a large case–control study using electronic records, Br. J. Cancer 108 (1) (2013) 25.
[18] A. Taylor, S. Stapley, W. Hamilton, Jaundice in primary care: a cohort study of adults aged> 45 years using electronic medical records, Fam. Pract. 29 (4) (2012) 416–420.
[19] S.J. Price, E.A. Shephard, S.A. Stapley, K. Barraclough, W.T. Hamilton, Non-visible versus visible haematuria and bladder cancer risk: a study of electronic records in primary care, Br. J. Gen. Pract. 64 (626) (2014) e584–e589.
[20] N.U. Din, O.C. Ukoumunne, G. Rubin, W. Hamilton, B. Carter, S. Stapley, R.D. Neal, Age and gender variations in cancer diagnostic intervals in 15 cancers: analysis of data from the UK Clinical Practice Research Datalink, PloS One 10 (5) (2015), e0127717.
[21] L.T. Mounce, S. Price, J.M. Valderas, W. Hamilton, Comorbid conditions delay diagnosis of colorectal cancer: a cohort study using electronic primary care records, Br. J. Cancer 116 (12) (2017) 1536–1543.
[22] S. Walker, C. Hyde, W. Hamilton, Risk of breast cancer in symptomatic women in primary care: a case–control study using electronic records, Br. J. Gen. Pr. 64 (629) (2014) e788–e793.
[23] N.O.-G.C. Audit, National Oesophago-Gastric Cancer Audit 2021, 2021 ⟨https://www.nogca.org.uk/reports/2021-annual-report/⟩.
[24] J.M. Valderas, B. Starfield, B. Sibbald, C. Salisbury, M. Roland, Defining comorbidity: implications for understanding health and health services, Ann. Fam. Med. 7 (4) (2009) 357–363.
[25] Y. Zhou, F.M. Walter, L. Mounce, G.A. Abel, H. Singh, W. Hamilton, G.D. Stewart, G. Lyratzopoulos, Identifying opportunities for timely diagnosis of bladder and renal cancer via abnormal blood tests: a longitudinal linked data study, Br. J. Gen. Pract. 72 (714) (2022) e19–e25.
[26] M. Moullet, G. Funston, L.T. Mounce, G.A. Abel, N. de Wit, F.M. Walter, Y. Zhou, Pre-diagnostic clinical features and blood tests in patients with colorectal cancer: a retrospective linked-data study, Br. J. Gen. Pract. 72 (721) (2022) e556–e563.
[27] S.W. Price, Bianca; Mounce, Luke; Hamilton, Willie; Abel, Gary, Example Stata syntax and data construction for negative binomial time series regression, Mendeley Data, 2022.
[28] T.K. Mukhtar, C. Bankhead, S. Stevens, R. Perera, T.A. Holt, C. Salisbury, F. R. Hobbs, Factors associated with consultation rates in general practice in England, 2013–2014: a cross-sectional study, Br. J. Gen. Pract. 68 (670) (2018) e370–e377.
[29] M.L. Buis, Stata tip 87: Interpretation of interactions in nonlinear models, Stata J. 10 (2) (2010) 305–308.
[30] J. Watson, B.D. Nicholson, W. Hamilton, S. Price, Identifying clinical features in primary care electronic health record studies: methods for codelist development, BMJ Open 7 (11) (2017), e019637.
[31] S.J. Price, S.A. Stapley, E. Shephard, K. Barraclough, W.T. Hamilton, Is omission of free text records a possible source of data loss and bias in Clinical Practice Research Datalink studies? A case-control study, BMJ Open 6 (5) (2016), e011664.
[32] R.A. Payne, S.C. Mendonca, M.N. Elliott, C.L. Saunders, D.A. Edwards, M. Marshall, M. Roland, Development and validation of the Cambridge Multimorbidity Score, CMAJ: Can. Med. Assoc. J. = J. De. l'Assoc. Med. Can. 192 (5) (2020). E107-e114.
[33] A. Cassell, D. Edwards, A. Harshfield, K. Rhodes, J. Brimicombe, R. Payne, S. Griffin, The epidemiology of multimorbidity in primary care: a retrospective cohort study, Br. J. Gen. Pract.: J. R. Coll. Gen. Pract. 68 (669) (2018) e245–e251.
[34] C.S. Arhi, A. Bottle, E.M. Burns, J.M. Clarke, P. Aylin, P. Ziprin, A. Darzi, Comparison of cancer diagnosis recording between the Clinical Practice Research Datalink, Cancer Registry and Hospital Episodes Statistics, Cancer Epidemiol. 57 (2018) 148–157.
[35] A. Dregan, H. Moller, T. Murray-Thomas, M.C. Gulliford, Validity of cancer diagnosis in a primary care database compared with linked cancer registrations in England. Population-based cohort study, Cancer Epidemiol. 36 (5) (2012) 425–429.