Prioritising primary care patients with unexpected weight loss for cancer investigation: diagnostic accuracy study

Brian D Nicholson,1 Paul Aveyard,1 Sarah J Price,2 FD Richard Hobbs,1 Constantinos Koshiaris,1 Willie Hamilton2

Objective

To quantify the predictive value of unexpected weight loss (WL) for cancer according to patient’s age, sex, smoking status, and concurrent clinical features (symptoms, signs, and abnormal blood test results).

Design

Diagnostic accuracy study.

Setting

Clinical Practice Research Datalink electronic health records data linked to the National Cancer Registration and Analysis Service in primary care, England.

Participants

63 973 adults (≤18 years) with a code for unexpected WL from 1 January 2000 to 31 December 2012.

Main outcome measures

Cancer diagnosis in the six months after the earliest weight loss code (index date). Codes for additional clinical features were identified in the three months before to one month after the index date. Diagnostic accuracy measures included positive and negative likelihood ratios, positive predictive values, and diagnostic odds ratios.

Results

Of 63 973 adults with unexpected WL, 37 215 (58.2%) were women, 33 167 (51.8%) were aged 60 years or older, and 16 793 (26.3%) were ever smokers. 908 (1.4%) had a diagnosis of cancer within six months of the index date, of whom 882 (97.1%) were aged 50 years or older. The positive predictive value for cancer was above the 3% threshold recommended by the National Institute for Health and Care Excellence for urgent investigation in male ever smokers aged 50 years or older, but not in women at any age. Additional clinical features were associated with cancer in men with unexpected WL, and 11 in women. Positive likelihood ratios in men ranged from 1.86 (95% confidence interval 1.32 to 2.62) for non-cardiac chest pain to 6.10 (3.44 to 10.79) for abdominal blood loss test results associated with cancer included low albumin levels (4.67, 4.14 to 5.38), calcium (4.28, 3.05 to 6.02), total white cell count (3.76, 3.30 to 4.28), and C reactive protein (3.59, 3.31 to 3.89). However, no normal blood test result in isolation ruled out cancer.

Conclusion

The risk of cancer in adults with unexpected WL presenting to primary care is 2% or less and does not merit investigation under current UK guidelines. However, in male ever smokers aged 50 years or older and in patients with concurrent clinical features, the risk of cancer warrants referral for invasive investigation. Clinical features typically associated with specific cancer sites are markers of several cancer types when they occur with unexpected WL.

Introduction

Unexpected weight loss (WL) is recorded for about 1.5% of adults attending primary care. The likelihood of a cancer diagnosis in such people is increased in the three to six months after the first record of unexpected WL compared with people without unexpected WL: men with unexpected WL are three times as likely as men without unexpected WL to have a diagnosis of cancer within three months and are twice as likely to receive a diagnosis within six months; women with unexpected WL are twice as likely to have a diagnosis of cancer within three months (see table 1). Both early and late stage cancers are associated with unexpected WL. The greatest risks are from lymphoma, cancer of unknown primary, or cancers of the pancreas, gastro-oesophageal tract, lung, bowel, or renal tract. A cancer diagnosis is less likely than in people without recorded unexpected WL after the initial three to six month period.
Table 1 | Number of adults (≥18 years old) with and without unexpected weight loss (WL) and outcome of cancer diagnosis within six months, by sex

| Weight loss status | Cancer (n=3019) | No cancer (n=327425) | Sensitivity (%) | Specificity (%) | PPV (%) |
|--------------------|----------------|----------------------|-----------------|----------------|--------|
| **All**            |                |                      |                 |                |        |
| Unexpected WL      | 908            | 63065                | 30.08           | 80.74          | 1.42   |
| No unexpected WL   | 2111           | 264360               |                 |                | 0.79   |
| **Men**            |                |                      |                 |                |        |
| Unexpected WL      | 548            | 26210                | 35.02           | 81.07          | 2.05   |
| No unexpected WL   | 1017           | 112277               |                 |                | 0.90   |
| **Women**          |                |                      |                 |                |        |
| Unexpected WL      | 360            | 36855                | 24.76           | 80.5           | 0.97   |
| No unexpected WL   | 1094           | 152133               |                 |                | 0.71   |

PPV=positive predictive value.
See Nicholson et al.2020 for description of matched cohort of five people with no unexpected WL to one person with unexpected WL.

Unexpected WL can also be caused by a wide range of benign and serious conditions associated with various bodily systems, lifestyle choices, and socioeconomic factors.1 Differential diagnoses include advanced heart failure, chronic obstructive pulmonary disease, renal disease, pancreatic insufficiency, malabsorption, and endocrine disease, with up to 25% of patients without a diagnosis to explain their weight loss after extended follow-up.3 4 The non-specific nature of unexpected WL creates the clinical problem of who should be investigated further for cancer—and possibly using invasive methods—and who could be spared investigation. Several clinical reviews have proposed plausible approaches assessing the risk of cancer, but evidence generally has been from studies of older people admitted to hospital for investigation.5 Such research does not directly help general practitioners to plan investigations in primary care because of spectrum bias.7 Given the absence of appropriate clinical guidelines, or standardised practice, doctors have been reported to take diverse action, from doing nothing to ordering “extensive blind investigations” because of the fear of underlying cancer.8 9

Most research on the predictive value of cancer related unexpected WL in primary care has included patients based on their final cancer diagnosis rather than on the weight loss.5 The evidence base informed the National Institute for Health and Care Excellence guidance on suspected cancer, which recommended further investigations for patients with a positive predictive value (PPV) for cancer exceeding 3%.10 Studies have investigated unexpected WL together with other symptoms and signs occurring over a one to two year period preceding the cancer diagnosis without acknowledging that the predictive value of individual symptoms will vary during different periods.5 11 12 In this context, predictive values could have been reported for pairs of clinical features that occurred months or years apart, potentially with different causes unrelated to the eventual cancer diagnosis.

Although simple blood tests are often used to investigate non-specific symptoms in primary care patients,13 14 the role of such tests in selecting those with unexpected WL for further cancer investigation is poorly understood. Abnormal test results might facilitate patient triage,15 16 be poor predictors of cancer,17 18 or be predictive across several cancer sites.20 Triage testing in primary care is important to avoid unnecessary urgent referrals of patients for invasive investigation.

We conducted a diagnostic accuracy study using routinely collected electronic health records in primary care to establish the predictive value of unexpected WL for cancer, given the patient’s age, sex, smoking status, concurrent symptoms, signs, and blood test results. To identify malignancies that might be prioritised for further investigation after referral, we considered the predictive value for cancer overall and by cancer site.

**Methods**

**Study design and population**

We used electronic health records from the Clinical Practice Research Datalink (CPRD), a representative database of anonymised primary care records covering 6.9% of the UK population,21 linked to the National Cancer Registrations and Analysis Service (NCRAS) cancer register. The “Performance of diagnostic strategies” section of the published protocol pertains to this analysis.9 We followed the RECORD (REporting of studies Conducted using Observational Routinely-collected Data) reporting guidelines.22 Study entry was from 1 January 2000 to 31 December 2012 to allow two years or more to accommodate the time it takes for NCRAS to release validated data.

Patients were included if they were aged 18 years or older, registered with a CPRD general practice, eligible for linkage to NCRAS and Office for National Statistics data and the index of multiple deprivation, and had at least one code for unexpected WL and at least 12 months of data before the first recorded unexpected WL code (the index date). These unexpected WL codes equated to a mean weight loss of 5% or more within a six month period in our previous internal validation study of weight related coding in CPRD.2 Unexpected WL could be coded according to a range of clinical scenarios, including unexpected WL reported as the patient’s presenting condition, after targeted history taking, and after weight measurement as part of the clinical examination or as part of a routine health check or chronic disease review.

We excluded patients if they had a prescription of weight reducing treatment (orlistat) or a code for bariatric surgery in the previous six months, or if they had a cancer diagnosis before the index date.
Cancer (reference standard)
To identify cancers, we updated an existing library of codes to include all high level ICD-0 (international classification of diseases for oncology) categories.\(^1\) Cancers classified as non-melanoma skin, in situ, benign, ill defined, or uncertain were excluded. Furthermore, we grouped cancer sites that would usually be investigated using the same test or by the same specialty—for example, renal, ureteric, and bladder as renal tract cancers, and liver, gallbladder, and biliary tree as hepatobiliary cancers. All cancers diagnosed in the six months after the index date were identified in the CPRD and linked NCRAS data. We used the first site specific cancer code after the index date to define cancer site. Cancer of unknown primary was defined if a code identifying a secondary cancer (such as lymph node metastasis or cerebral metastasis) was found but there was no code for a primary cancer.

Sociodemographic and clinical features
Sociodemographic details coded on or before the index date were extracted from the CPRD records. We identified codes related to signs and symptoms and blood test results in the three months before to one month after the index date. A long list of symptoms and signs shown to have an independent association with undiagnosed cancer were selected either through their inclusion in the NICE NG12 guidance for suspected cancer or based on studies published after the NICE guidance (ie, central nervous system malignancies and head and neck cancers) (supplementary appendix 1). For blood tests, we identified those most commonly requested within the four month period, dropped outliers and erroneous results, and dichotomised continuous test results as abnormal or normal using standard laboratory ranges (supplementary appendix 2).

Box 1: Classification of true and false positive and negative test results

**True positive result**
- Presence of a clinical feature recorded in Clinical Practice Research Datalink (CPRD) in the three months before to one month after the index date of unexpected weight loss (WL) in patients with a cancer diagnosis (recorded in CPRD or National Cancer Registrations and Analysis Service (NCRAS) cancer registry) in the six months after the index date

**False positive result**
- Presence of a clinical feature recorded in CPRD in the three months before to one month after the index date of unexpected WL in patients with no cancer diagnosis (recorded in CPRD or NCRAS cancer registry) in the six months after the index date

**False negative result**
- Absence of a clinical feature recorded in CPRD in the three months before to one month after the index date of unexpected WL in patients with a cancer diagnosis (recorded in CPRD or NCRAS cancer registry) in the six months after the index date

**True negative result**
- Absence of a clinical feature recorded in CPRD in the three months before to one month after the index date of unexpected WL in patients with no cancer diagnosis (recorded in CPRD or NCRAS cancer registry) in the six months after the index date

Statistical analysis
Box 1 shows the definitions of true positive, false positive, false negative, and true negative test results for clinical features. For combinations of unexpected WL and age group, sex, smoking status, clinical features, and abnormal blood test results, we estimated diagnostic accuracy statistics for the cancer outcome using 2×2 tables with the DIAGT Stata module: positive likelihood ratios, negative likelihood ratios, positive predictive values, and diagnostic odds ratios along with 95% confidence intervals. A rule of thumb is that a test with a positive likelihood ratio of 5 or more is good for ruling in disease and a test with a negative likelihood ratio of 0.2 or less is good for ruling out disease.\(^23\) 25 The analysis was conducted for cancer overall and by cancer site. To select symptoms and signs, we used multivariable backwards stepwise logistic regression starting with all symptoms, signs, and sociodemographic factors as independent covariates, using a P value of 0.05 or less for retention (supplementary appendix 1). We elected to use an indicator variable over multiple imputation to replace missing data on lifestyle factors, as the main purpose of including the covariates was to reduce confounding in variable selection rather than to identify the association between the lifestyle covariate with missing data and cancer.

In discrete analyses we also calculated diagnostic accuracy statistics for each of the 10 most commonly recorded blood tests, and included only patients with each blood test result. When tests were components of another test, we chose the quantum—for example, using total white cell count rather than white cell subtypes.

Sensitivity analysis
We repeated the selection process for clinical features using an interval of three months before the index date to the first date of either three months after or the cancer diagnosis to explore whether broadening the window for inclusion of clinical features changed our findings.

Patient and public involvement
Patients and the public were involved in an advisory capacity in the application for funding to support this research. An advisory panel of patient and public members provided comments on a related article that informed the current analysis. Patients were not directly involved in the conduct or analysis of the study. The results of this study will be disseminated through the media channels of the host institution of the lead author and the funder for scientific and lay audiences.

Results
Figure 1 shows the flow of participants through the trial. Of 63 973 patients with a code for unexpected WL, 37 215 (58.2%) were women and 26 758 (41.8%) were men; 33 156 (51.8%) were aged 60 years or older, 33 846 (52.9%) had a normal body mass index, and 16 793 (26.3%) were ever smokers (table 2). The most
common features to be recorded alongside unexpected WL were cough (7.6%), abdominal pain (5.9%), back pain (5.1%), chest infection (4.7%), and fatigue (4.5%) (supplementary appendix 3). The most commonly recorded tests were full blood count (predominantly for haemoglobin (72.1%), platelets (70.7%), total white cell count (69.9%), and creatinine (69.8%)), and liver function tests—namely, bilirubin (65.5%), albumin (65.1%), and alkaline phosphatase (64.5%) (supplementary appendix 2).

Age, sex, and smoking status
The positive predictive value for a cancer diagnosis was higher in people who were older and those who smoked (fig 2)—it was more than 2% in patients aged 50 years or older. Analysis by sex, however, showed that the positive predictive value was more than 3% for male ever smokers aged 50 years or older, whereas the positive predictive value remained less than 2% for women, except in smokers aged 70 years or older.

Signs and symptoms
In multivariable analysis, features selected to be positively associated with cancer in people with unexpected WL were abdominal pain, appetite loss, abdominal mass, iron deficiency anaemia, jaundice, chest signs, and lymphadenopathy (table 3). Dysphagia, haemoptysis, and non-cardiac chest pain were associated with cancer only in men with unexpected WL, and back pain, change in bowel habit, dyspepsia, and venous thromboembolism only in women with unexpected WL. Positive likelihood ratios in men ranged from 1.32 (95% confidence interval 1.27 to 1.45) to 6.10 (3.44 to 10.79) for abdominal mass, and in women from 1.62 (1.51 to 2.29) for back pain to 20.9 (10.7 to 40.9) for jaundice. Although four symptoms and signs had positive likelihood ratios greater than 5, they were relatively uncommon, each occurring in six to 10 people with a diagnosis of cancer, depending on sex. No negative likelihood ratio was below 0.2, with values ranging from 0.94 to 1.00 (table 3).

For men and women aged 60 years or older, unexpected WL and the co-occurrence of the selected symptoms and signs were associated with an increase in the positive predictive value by 3% or higher and above the underlying positive predictive value for all people with unexpected WL in each age group (first row of fig 3 for women and fig 4 for men). For men aged

Table 2 | Baseline characteristics of study population

| Characteristics     | Men (n=26 758) | Women (n=37 215) |
|---------------------|----------------|------------------|
| Age (years):        |                |                  |
| 18-39               | 14 290 (22.3)  | 13 418 (36.1)    |
| 40-49               | 8016 (12.5)    | 7 146 (19.1)     |
| 50-59               | 8 511 (13.3)   | 7 504 (19.9)     |
| 60-69               | 9 017 (14.1)   | 8 487 (22.7)     |
| 70-79               | 11 565 (18.1)  | 10 059 (27.1)    |
| ≥80                 | 12 574 (19.7)  | 11 636 (31.4)    |
| Smoking status:     |                |                  |
| Current smoker      | 9629 (15.1)    | 7 397 (20.0)     |
| Former smoker       | 7 164 (11.1)   | 6 613 (17.7)     |
| Non-smoker          | 14 457 (22.6)  | 13 205 (35.3)    |
| Missing             | 32 723 (51.2)  | 24 091 (65.0)    |
| Alcohol intake status: |          |                  |
| Current drinker     | 18 435 (28.8)  | 14 940 (40.5)    |
| Non-drinker         | 8 095 (12.7)   | 6 637 (18.0)     |
| Former drinker      | 10 871 (17.2)  | 9 346 (25.5)     |
| Missing             | 36 356 (56.8)  | 28 403 (78.5)    |
| Body mass index:    |                |                  |
| Underweight         | 6 691 (10.9)   | 5 679 (15.5)     |
| Normal              | 33 846 (52.9)  | 28 540 (76.6)    |
| Overweight          | 10 790 (16.9)  | 9 341 (25.6)     |
| Obese               | 5 141 (8.0)    | 4 122 (11.4)     |
| Missing             | 7 235 (11.3)   | 6 503 (17.7)     |
| Cancer diagnosis:   |                |                  |
| Yes                 | 908 (1.4)      | 706 (1.9)        |
| No                  | 63 065 (98.6)  | 37 509 (98.1)    |

Fig 1 | Flow chart of participants through trial. NHS=National Health Service; CPRD=Clinical Practice Research Datalink; NCRAS=National Cancer Registrations and Analysis Service; ONS=Office for National Statistics; IMD=index of multiple deprivation

Source population: NHS patients aged ≥18 years registered with an up-to-standard general practice with acceptable CPRD records with eligibility for linkage to NCRAS, ONS, and IMD between 1 January to 31 December 2012

Patients with at least one code for unexpected weight loss

Excluded

6 220 Previous cancer
7 37 Recent weight loss intervention

Index test: first weight loss code

Reference standard: cancer diagnosis in CPRD or NCRAS within 6 months
40 to 59 years with unexpected WL, the co-occurrence of jaundice and lymphadenopathy was also associated with an increase in the positive predictive value of 3% or higher, whereas only abdominal mass reached this level for women in this age group (fig 3 for women and fig 4 for men).

**Blood tests**

Several abnormal blood test results in combination with unexpected WL showed the highest positive likelihood ratio values, some with upper 95% confidence intervals greater than 5: low albumin (4.67 (95% confidence interval 4.14 to 5.27), raised platelet levels (4.57 (3.88 to 5.38)), raised calcium levels (4.28 (3.05 to 6.02)), raised total white cell counts (3.76 (3.30 to 4.28)), and raised C reactive protein levels (3.59 (3.11 to 3.89)) (table 3). Normal inflammatory markers had the lowest negative likelihood ratios: C reactive protein (0.35 (95% confidence interval 0.29 to 0.43)) and erythrocyte sedimentation rate (0.42 (0.36 to 0.49)). Individual blood tests therefore did not reach the ideal threshold of 5 to rule in a cancer diagnosis or to 0.2 to rule out a diagnosis.

For example, women with dyspepsia and unexpected WL, they were associated with several cancer types. However, when they co-occurred with unexpected WL (fig 3 for women and fig 4 for men), because not all patients had all tests (fig 5).

**Cancer diagnoses**

Cancer was diagnosed in 908 (1.4%) patients within six months of the index date, of whom 882 (97.1%) were aged 50 years or older and 902 (99.3%) were aged 40 years or older. The most commonly diagnosed malignancies were cancers of the lung (n=220, 24.2%), bowel (114, 12.6%), gastro-oesophagus (103, 11.3%), and pancreas (80, 8.81%), and lymphoma (68, 7.49%).

Individually, some clinical features are generally considered to be associated with a single cancer site; however, when they co-occurred with unexpected WL, they were associated with several cancer types. For example, women with dyspepsia and unexpected WL were diagnosed as having cancers of the following types or site (in rank order): stomach or oesophagus,
bowl, pancreas, lung, bone connective or soft tissues, lymphoma, unknown primary, other, breast, central nervous system, and leukaemia (fig 3).

Similarly, abnormal test results in patients with unexpected WL were also associated with multiple cancer sites. For example, women aged 60-79 years with low albumin levels were diagnosed as having cancers of the following types or site (in rank order): lung, lymphoma, bowel, unknown primary, renal tract, stomach or oesophagus, ovary, hepatobiiliary, uterus, other, pancreas, central nervous system, breast, myeloma, and leukaemia (fig 3).

Isolated unexpected WL
Positive predictive values for patients without any of the selected clinical features were lower across every age range compared with the full cohort (fig 2). In addition, 13,941 (21.8%) patients had no record of a blood test, 142 (1.0%) of whom had a diagnosis of cancer. Overall, 89 patients with unexpected WL and cancer had neither a clinical feature nor blood test on record: 52 out of 57 men and 27 of 32 women were aged 60 years or older.

Sensitivity analyses
Appendix 3 shows the results of the sensitivity analysis. Widening the time window made almost no difference to the clinical features selected for inclusion in the final analysis.

Clinical guideline
Appendix 4 summarises the current NICE guideline recommendations for suspected cancer in patients with unexpected WL. Table 4 outlines an updated clinical guideline based on the results of this analysis.

---

**Table 3 | Predictive values for cancer over six months by clinical features and blood tests in adults aged 18 years or older attending primary care with unexpected weight loss**

| Clinical features | True positive | False positive | False negative | True negative | PLR (95% CI) | NLR (95% CI) | DOR (95% CI) |
|------------------|---------------|----------------|----------------|---------------|--------------|--------------|--------------|
| **Men**          |               |                |                |               |              |              |              |
| Abdominal pain   | 61            | 1336           | 487            | 24,874        | 2.18 (1.71 to 2.78) | 0.94 (0.91 to 0.96) | 2.33 (1.78 to 3.06) |
| Appetite loss    | 38            | 615            | 510            | 25,595        | 2.96 (2.15 to 4.06) | 0.95 (0.93 to 0.98) | 3.10 (2.21 to 4.35) |
| Dysphagia        | 21            | 290            | 527            | 25,920        | 3.46 (2.24 to 5.35) | 0.97 (0.96 to 0.99) | 3.56 (2.28 to 5.57) |
| Haemoptysis      | 9             | 101            | 539            | 26,109        | 4.26 (2.17 to 8.38) | 0.99 (0.98 to 1.00) | 4.32 (2.20 to 8.47) |
| Non-cardiac chest pain | 32  | 823            | 516            | 25,387        | 1.86 (1.32 to 2.62) | 0.97 (0.95 to 0.99) | 1.91 (1.33 to 2.75) |

**Women**

| Abdominal mass | 13            | 102            | 535            | 26,108        | 6.10 (3.44 to 10.79) | 0.98 (0.97 to 0.99) | 6.22 (3.50 to 11.1) |
| Chest signs    | 6             | 53             | 542            | 26,157        | 5.41 (3.24 to 12.25) | 0.99 (0.98 to 1.00) | 5.46 (2.39 to 12.5) |
| Iron deficiency anaemia | 22  | 212            | 526            | 25,998        | 4.96 (3.23 to 7.64) | 0.97 (0.95 to 0.98) | 5.13 (3.29 to 8.00) |
| Jaundice       | 7             | 59             | 541            | 26,151        | 5.67 (2.60 to 12.4) | 0.99 (0.98 to 1.00) | 5.74 (2.66 to 12.4) |
| Lymphadenopathy| 4             | 73             | 544            | 26,137        | 2.62 (0.96 to 7.14) | 1.00 (0.99 to 1.00) | 2.63 (1.00 to 6.96) |

**Liver function tests:**

| Low albumin    | 202           | 2665           | 463            | 38,292        | 4.67 (4.14 to 5.27) | 0.74 (0.71 to 0.78) | 6.27 (5.93 to 7.43) |
| Raised alkaline phosphatase | 271 | 6753           | 337            | 38,882        | 2.54 (2.11 to 2.78) | 0.69 (0.65 to 0.74) | 3.66 (3.12 to 4.28) |
| Raised bilirubin| 47            | 2028           | 598            | 39,222        | 1.48 (1.12 to 1.96) | 0.98 (0.95 to 1.00) | 1.52 (1.13 to 2.05) |

**Full blood count:**

| Low haemoglobin | 341           | 7099           | 374            | 38,315        | 3.05 (2.82 to 3.30) | 0.62 (0.58 to 0.67) | 4.92 (4.24 to 5.71) |
| Raised total white cell count | 179  | 3066           | 505            | 40,964        | 3.76 (3.30 to 4.28) | 0.79 (0.76 to 0.83) | 4.74 (3.98 to 5.64) |
| Raised platelets | 127           | 1776           | 570            | 42,766        | 4.57 (3.88 to 5.38) | 0.85 (0.82 to 0.88) | 5.37 (4.60 to 6.54) |

**Inflammatory markers:**

| Raised C reactive protein | 194 | 2891           | 76             | 11,542        | 3.59 (3.31 to 3.89) | 0.35 (0.29 to 0.43) | 10.2 (7.80 to 13.3) |
| Raised erythrocyte sedimentation rate | 274  | 6511           | 116            | 16,030        | 2.43 (2.27 to 2.60) | 0.42 (0.36 to 0.49) | 5.80 (4.67 to 7.26) |

**Biochemistry:**

| Raised calcium | 31            | 398            | 351            | 19,432        | 4.28 (3.05 to 6.02) | 0.93 (0.90 to 0.96) | 4.59 (3.18 to 6.64) |
| Raised creatinine | 254           | 12971          | 445            | 31,001        | 1.23 (1.12 to 1.36) | 0.90 (0.85 to 0.96) | 1.36 (1.17 to 1.59) |
| UWL plus | PPV (95% CI) by age group (years) | Cancer sites from highest to lowest risk |
|----------|----------------------------------|----------------------------------------|
| All women with UWL: 0.97 (0.87 to 1.07) 0.29 (0.19 to 0.43) 1.61 (1.38 to 1.86) 1.83 (1.56 to 2.13) | |  |
| Symptoms | | |
| Abdominal pain | 2.20 (1.65 to 2.87) 1.22 (0.49 to 2.50) 3.99 (2.69 to 5.68) 3.55 (2.00 to 5.79) | Pancreas, bowel, hepatobiliary, cancer of unknown primary, gastro-oesophageal, lymphoma, lung, ovary, central nervous system, renal tract, myeloma, uterine, leukemia, other |
| Appetite loss | 2.45 (1.56 to 3.65) 0.00 (0.00 to 2.59) 3.43 (1.72 to 6.05) 3.69 (1.86 to 6.51) | Bowel, lung, pancreas, gastro-oesophageal, head and neck, hepatobiliary, cancer of unknown primary, lymphoma, other |
| Back pain | 1.56 (1.05 to 2.27) 0.88 (0.24 to 2.23) 2.54 (1.51 to 3.99) 1.64 (0.64 to 3.35) | Lung, pancreas, cancer of unknown primary, renal tract, bowel, gastro-oesophageal, breast, myeloma, other, hepatobiliary, bowel connective and soft tissue, ovary, lymphoma |
| Change in bowel habit | 3.34 (1.74 to 5.77) 1.10 (0.03 to 5.97) 2.86 (0.78 to 7.19) 7.37 (3.01 to 14.59) | Bowel, lymphoma, pancreas, hepatobiliary, gastro-oesophageal, renal tract, ovary, uterine |
| Dyspepsia | 2.43 (1.58 to 3.57) 1.06 (0.22 to 3.07) 3.21 (1.67 to 5.54) 4.05 (1.87 to 7.58) | Gastro-oesophageal, bowel, pancreas, lung, bowel connective and soft tissue, lymphoma, cancer of unknown primary, other, breast, central nervous system, leukemia |
| Signs | | |
| Abdominal mass | 5.08 (1.89 to 10.76) 3.03 (0.08 to 15.76) 8.89 (2.48 to 21.22) 2.86 (0.07 to 14.92) | Bowel, gastro-oesophageal, lung, head and neck, renal tract |
| Chest signs | 10.34 (2.19 to 27.39) 0.00 (0.00 to 52.18) 10.00 (0.25 to 44.50) 14.29 (1.78 to 42.81) | Lung, bowel, cancer of unknown primary |
| Iron deficiency anaemia | 3.49 (2.11 to 5.39) 1.83 (0.12 to 4.67) 5.70 (2.64 to 10.54) 4.02 (1.63 to 8.11) | Bowel, gastro-oesophageal, lung, renal tract, bowel connective and soft tissue, uterine, ovary, cancer of unknown primary, other |
| Jaundice | 16.95 (8.44 to 32.89) 0.00 (0.00 to 21.81) 33.30 (14.59 to 56.97) 15.00 (3.27 to 37.89) | Hepatobiliary, pancreas, bowel, gastro-oesophageal |
| Lymphadenopathy | 2.27 (0.47 to 6.50) 0.00 (0.00 to 10.28) 0.00 (0.00 to 14.83) 27.27 (6.02 to 60.97) | Lymphoma |
| Venous thromboembolism | 5.88 (2.40 to 11.74) 0.00 (0.00 to 18.53) 8.00 (2.22 to 19.23) 6.98 (1.46 to 19.06) | Lung, cancer of unknown primary, pancreas, lymphoma, ovary |
| Liver function tests | | |
| Albumin (low) | 4.79 (1.89 to 10.76) 1.83 (0.38 to 5.25) 8.91 (65.53 to 11.77) 3.71 (0.04 to 0.46) | Lung, lymphoma, bowel, cancer of unknown primary, renal tract, gastro-oesophageal, ovary, hepatobiliary, uterine, other, pancreas, central nervous system, breast, myeloma, leukemia |
| Alkaline phosphatase (raised) | 2.65 (2.21 to 3.15) 1.07 (0.49 to 2.03) 3.35 (2.60 to 4.23) 3.12 (2.31 to 4.10) | Lung, bowel, lymphoma, gastro-oesophageal, pancreas, hepatobiliary, cancer of unknown primary, renal tract, central nervous system, ovary, uterine, other, breast, melanoma, bowel connective and soft tissue, head and neck, myeloma |
| Bilirubin (raised) | 1.85 (1.01 to 3.08) 0.00 (0.00 to 2.07) 4.64 (2.34 to 8.15) 1.83 (0.38 to 5.25) | Hepatobiliary, pancreas, lung, gastro-oesophageal, bowel, breast |
| Full blood count | | |
| Haemoglobin (low) | 1.69 (1.47 to 1.95) 1.52 (0.56 to 3.27) 4.16 (3.03 to 5.57) 3.71 (2.32 to 4.21) | Lung, bowel, lymphoma, gastro-oesophageal, cancer of unknown primary, renal tract, hepatobiliary, pancreas, ovary, uterine, breast, bowel connective and soft tissue, central nervous system, myeloma, leukemia, other, head and neck |
| Total white cell count (raised) | 4.48 (3.54 to 5.32) 2.49 (1.05 to 4.30) 5.96 (4.18 to 7.95) 6.92 (4.75 to 9.68) | Lung, bowel, cancer of unknown primary, gastro-oesophageal, pancreas, lymphoma, hepatobiliary, renal tract, uterine, ovary, breast, other, bowel connective and soft tissue, head and neck |
| Platelets (raised) | 4.66 (3.54 to 6.00) 3.08 (1.14 to 6.58) 6.41 (4.53 to 8.76) 3.72 (2.00 to 6.29) | Bowel, lung, renal tract, cancer of unknown primary, gastro-oesophageal, hepatitis, gastro-oesophageal, ovary, lymphoma, other, pancreas, bowel connective and soft tissue, breast, uterine |
| Inflammatory markers | | |
| C reactive protein (raised) | 4.83 (1.81 to 6.02) 2.00 (0.55 to 5.04) 5.58 (3.98 to 7.58) 5.78 (3.96 to 8.11) | Lung, cancer of unknown primary, lymphoma, bowel, gastro-oesophageal, renal tract, pancreas, ovary, uterine, breast, hepatobiliary, bowel connective and soft tissue, central nervous system, melanoma, head and neck, leukemia |
| Erythrocyte sedimentation rate (raised) | 2.94 (2.40 to 3.53) 1.24 (0.50 to 2.53) 4.10 (1.10 to 5.30) 2.91 (1.97 to 4.13) | Lung, lymphoma, bowel, gastro-oesophageal, cancer of unknown primary, pancreas, renal tract, hepatobiliary, ovary, uterine, other, central nervous system, breast, head and neck, central nervous system, melanoma, leukemia |
| Biochemistry | | |
| Calcium (raised) | 4.50 (2.93 to 6.57) 2.50 (0.06 to 13.16) 8.63 (4.54 to 14.59) 4.90 (1.81 to 11.07) | Lung, renal tract, lymphoma, cancer of unknown primary, bowel, bowel connective and soft tissue, myeloma |
| Creatinine (raised) | 1.33 (1.11 to 1.58) 0.15 (0.02 to 0.53) 1.74 (1.32 to 2.23) 1.79 (1.38 to 2.28) | Lung, bowel, renal tract, lymphoma, cancer of unknown primary, pancreas, gastro-oesophageal, hepatobiliary, ovary, other, uterine, breast, bowel connective and soft tissue, central nervous system, melanoma, myeloma, head and neck |

Fig 3 | Positive predictive values (PPVs) of symptoms, signs, and blood tests for a cancer diagnosis within six months in women with unexpected weight loss (UWL) by age group. Red shading represents a PPV of 3% or higher, the threshold above which the National Institute for Health and Care recommends investigation for cancer. Yellow shading represents a PPV of 2-3%.
| UWL plus | Age group (years) | Cancer sites from highest to lowest risk |
|----------|------------------|----------------------------------------|
|          | all men with UWL |                                        |
|          | ±18              | 4.05 (1.88 to 2.22) 0.56 (0.40 to 0.79) 3.22 (2.88 to 3.60) 4.58 (3.98 to 5.29) |
| Abdominal pain | 4.37 (3.36 to 5.57) | 2.02 (0.87 to 3.93) 7.03 (5.00 to 9.57) 7.58 (4.40 to 12.02) Bowen, pancreas, lung, cancer of unknown primary, lymphoma, gastro-oesophageal, renal tract, hepatobiliary, prostate, bowel connective and soft tissue |
| Appetite loss | 5.82 (4.15 to 7.90) | 2.14 (0.44 to 6.13) 8.40 (5.27 to 12.53) 8.19 (4.55 to 13.38) Lung, gastro-oesophageal, bowel, pancreas, renal tract, cancer of unknown primary, lymphoma, leukemia, bowel connective and soft tissue, myeloma, prostate |
| Dysphagia | 6.75 (4.23 to 10.14) | 1.59 (0.04 to 8.5) 10.42 (9.95 to 16.6) 6.17 (2.03 to 18.82) Gastro-oesophageal, lung, head and neck, bowel, myeloma |
| Jaundice | 8.18 (3.81 to 14.96) | 3.57 (0.09 to 18.3) 13.33 (5.05 to 26.7) 9.52 (1.17 to 30.3) Lung, lymphoma, cancer of unknown primary |
| Non-cardiac chest pain | 3.74 (2.57 to 5.24) | 0.79 (0.10 to 2.82) 6.19 (3.82 to 9.40) 7.30 (3.56 to 13.0) Lung, gastro-oesophageal, lymphoma, bowel, cancer of unknown primary, prostate, hepatobiliary, panneumia, myeloma, leukemia |
| Abdominal mass | 11.30 (6.16 to 18.5) | 7.69 (0.95 to 25.1) 9.62 (3.20 to 21.0) 20.00 (7.71 to 38.5) Bowen, gastro-oesophageal, lung, lymphoma, cancer of unknown primary |
| Chest signs | 10.17 (3.82 to 20.8) | 0.00 (0.00 to 45.9) 12.50 (3.51 to 28.9) 9.52 (1.17 to 30.3) Bowel connective and soft tissue, bowel, gastro-oesophageal, other |
| Iron deficiency anaemia | 9.40 (5.99 to 13.8) | 0.00 (0.00 to 12.7) 12.80 (7.50 to 19.9) 8.57 (3.21 to 17.3) Bowen, gastro-oesophageal, lung, renal tract, pancreas, hepatobiliary, lymphoma, bowel connective and soft tissue |
| Jaundice | 10.61 (4.37 to 20.6) | 11.76 (4.16 to 36.4) 7.41 (0.91 to 24.2) 30.00 (6.67 to 65.2) Pancreas, hepatobiliary, lung, bowel connective and soft tissue |
| Lymphadenopathy | 5.19 (1.43 to 12.7) | 8.33 (1.03 to 27.0) 7.14 (0.18 to 33.8) 16.67 (4.02 to 64.1) Lymphoma, renal tract, cancer of unknown primary |
| Albumin (low) | 9.45 (7.96 to 11.1) | 3.81 (1.66 to 7.7) 11.95 (9.53 to 14.7) 9.20 (6.84 to 12.0) Lung, bowel, gastro-oesophageal, renal tract, lymphoma, hepatobiliary, cancer of unknown primary, pancreas, prostate, leukemia, bowel connective and soft tissue, head and neck, central nervous system, myeloma |
| Alkaline phosphatase (raised) | 6.31 (5.35 to 7.38) | 3.01 (1.73 to 4.83) 8.25 (6.61 to 10.1) 8.30 (6.17 to 10.8) Lung, pancreas, cancer of unknown primary, hepatobiliary, bowel, lymphoma, prostate, renal tract, gastro-oesophageal, bowel connective and soft tissue, other, melanoma, myeloma, leukemia |
| Bilirubin (raised) | 2.50 (1.73 to 3.5) | 0.54 (0.07 to 1.93) 3.71 (2.21 to 5.8) 6.13 (3.31 to 10.2) Hepatobiliary, pancreas, lung, bowel, lymphoma, cancer of unknown primary, gastro-oesophageal, renal tract, prostate, leukemia |
| Full blood count | Haemoglobin (low) | 4.03 (3.61 to 4.49) 2.09 (0.25 to 3.7) 6.76 (5.68 to 7.97) 6.34 (5.21 to 7.83) Lung, lymphoma, bowel, gastro-oesophageal, renal tract, prostate, cancer of unknown primary, pancreas, hepatobiliary, bowel connective and soft tissue, myeloma, leukemia, head and neck, other |
| Total white cell count (raised) | 6.80 (5.57 to 8.21) | 2.49 (1.25 to 4.4) 9.85 (7.56 to 12.5) 11.65 (8.06 to 16.1) Lung, bowel, gastro-oesophageal, prostate, renal tract, lymphoma, pancreas, cancer of unknown primary, hepatobiliary, bowel connective and soft tissue, head and neck, leukemia |
| Platelets (raised) | 10.14 (8.01 to 12.6) | 4.71 (2.05 to 9.06) 12.57 (9.22 to 16.6) 15.33 (9.75 to 22.4) Lung, bowel, renal tract, lymphoma, gastro-oesophageal, cancer of unknown primary, prostate, pancreas, bowel connective and soft tissue, hepatobiliary, head and neck |
| Inflammatory markers | C reactive protein (raised) | 7.73 (6.45 to 9.17) 3.58 (1.80 to 6.32) 10.11 (8.00 to 12.5) 8.82 (6.26 to 12.0) Lung, lymphoma, renal tract, bowel, pancreas, cancer of unknown primary, gastro-oesophageal, hepatobiliary, prostate, bowel connective and soft tissue, head and neck |
| Erythrocyte sedimentation rate (raised) | 5.82 (5.00 to 6.72) | 0.83 (0.33 to 1.7) 7.01 (5.81 to 8.36) 6.94 (5.33 to 8.8) Lung, gastro-oesophageal, bowel, lymphoma, renal tract, prostate, cancer of unknown primary, pancreas, hepatobiliary, myeloma, bowel connective and soft tissue, head and neck, leukemia |
| Biochemistry | Calcium (raised) | 7.55 (4.74 to 11.3) 8.00 (0.98 to 26.0) 17.86 (8.91 to 30.4) 10.71 (2.27 to 28.2) Lung, lymphoma, renal tract, cancer of unknown primary, gastro-oesophageal, hepatobiliary, other |
| Creatinine (raised) | 3.34 (2.80 to 3.96) | 0.67 (0.14 to 1.94) 3.44 (2.64 to 4.41) 4.41 (3.43 to 5.5) Lung, prostate, gastro-oesophageal, renal tract, lymphoma, bowel, cancer of unknown primary, prostate, hepatobiliary, leukemia, bowel connective and soft tissue, myeloma, other |

Fig 4 | Positive predictive values (PPVs) of symptoms, signs, and blood tests for a cancer diagnosis within six months in men with unexpected weight loss (UWL) by age group. Red shading represents a PPV of 3% or higher, the threshold above which the National Institute of Health and Care Excellence recommends investigation for cancer. Yellow shading represents a PPV of 2-3%
Discussion
The risk of undiagnosed cancer in adults with recorded unexpected WL alone is below the UK’s current 3% threshold that warrants investigation. However, in male ever smokers aged 50 years or older and in all patients with concurrent clinical features (certain symptoms and signs or abnormal results for simple blood tests) the probability of undiagnosed cancer rises above 3%. These features are abdominal mass, abdominal pain, appetite loss, chest signs, iron deficiency anaemia, jaundice, and lymphadenopathy in both men and women; dysphagia, haemoptysis, and non-cardiac chest pain in men; and back pain, change in bowel habit, dyspepsia, and venous thromboembolism in women. The abnormal blood test results in men and women are low albumin levels and raised levels of white cell counts, calcium, platelets, and inflammatory markers. The absence of individual clinical features in the three months before and up to one month after the index date, or the presence of individual normal blood tests in this time window, does not confidently rule-out cancer in patients with unexpected WL.

Strengths and limitations of this study
We took the following steps to maximise the likelihood that the unexpected WL cohort was accurately defined. First we confirmed that insufficient weight measurements were recorded in UK primary care to define unexpected WL, with clustering of weight recording noted in women with higher body mass index and in those with comorbidity. We then conducted an internal validation study to identify which codes most consistently defined unexpected WL and investigated whether weight measurements could be used in preference to codes. We included each patient once in the analysis by choosing the first unexpected WL code and excluded patients with a history of cancer to ensure we were investigating unexpected WL associated with a first diagnosis of cancer. We excluded patients with objective evidence of deliberate weight loss (ie, prescription records and coding for bariatric surgery). The presence of advanced comorbid conditions might be more likely to be associated with unexpected WL and could modify the association of unexpected WL and cancer. As it is problematic to identify disease
Table 4 | Revised clinical recommendations for specialised investigation for cancer in patients with unexpected weight loss

| Age (years) | Men and women | Women |
|-------------|---------------|-------|
| ≥40         | Abdominal mass, thrombocytosis | Jaundice, lymphadenopathy, hypoalbuminaemia, hypercalcaemia, raised alkaline phosphatase level, raised C reactive protein level |
| ≥60         | Abdominal pain, anaemia, appetite loss, chest signs, hyperbilirubinaemia, leucocytosis | Dysphagia, haemoptysis, non-cardiac chest pain, raised erythrocyte sedimentation rate |
| ≥80         | Ever smokers | Dyspepsia, jaundice, venous thromboembolism, hypoalbuminaemia, hypercalcaemia, raised alkaline phosphatase level, raised C reactive protein level, raised erythrocyte sedimentation rate |
| ≥100        | Change in bowel habit, lymphadenopathy | |

severity using CPRD data, however, we did not exclude patients with these conditions.

We analysed clinical features as occurring with unexpected WL if they were coded in the three months before and the one month after the index date. This was a clinical decision, as the epidemiology on this topic is limited, based on consideration that a general practitioner is likely to look back at recent notes and investigate unexpected WL within a month of presentation. Symptoms occurring more than three months before the index date might well be unlinked to the unexpected WL. Some studies have reported the frequency of individual clinical features for cancer cases and controls before a cancer diagnosis, with few symptoms more common in cases than controls earlier than six months before the cancer diagnosis.17 25 26 None of these studies, however, reported the timing of multiple symptoms leading up to a cancer diagnosis. We have shown previously that cancers are likely to be diagnosed in patients within three to six months of presentation with unexpected WL, and our findings did not change significantly in sensitivity analysis extending the period to capture co-occurring symptoms and signs up to the day of cancer diagnosis.1 The high number of false negatives observed for individual symptoms and signs accompanying unexpected WL indicates the varied clinical presentation of cancer associated with unexpected WL. Accompanying symptoms and signs are only recorded if patients experience them, remember to report them, or they are uncovered by the doctor during a clinical examination.

We also conducted individual analyses for each blood test, including only patients with a result for that blood test. Previous studies have replaced missing blood tests with negative results to allow full case multivariable analysis. We decided against this for two reasons. Firstly, patients who have been tested represent a higher risk population than those who have not been tested.19 27 and therefore people with a normal test result might not have the same likelihood of undiagnosed cancer as people who have not been tested. It is unclear how this testing bias relates to the study participants with unexpected WL, for whom the blood test was taken close enough to the index date for us to be confident that the test and the result pertained to it. Secondly, classifying absent tests as negative results inflates the number of “true” negatives and misestimates diagnostic accuracy, making it difficult to interpret negative likelihood ratios and negative predictive values. However, as not all patients had been tested with all blood tests, we could not calculate precise estimates for combinations of multiple blood test results. We also dichotomised continuous blood test results at thresholds used in clinical practice to signify an abnormal result. By dichotomising we lose information by classifying markedly abnormal test results together with mildly abnormal results. The positive predictive values presented can therefore be considered conservative estimates of the associated cancer risk. The high number of false positive blood test results represent that doctors decided to use blood tests to investigate unexpected WL in most cases, that cancer was associated with unexpected WL in fewer than 2% of cases, and that abnormal results in the blood tests studied are not only found in cases of cancer. Finally, we classified people as having cancer if a code was entered within six months of presenting with unexpected WL. Previous research has shown that if cancer is not diagnosed within six months, the risk of cancer being the cause of the unexpected WL is low.1 28

Comparison with other studies

A 2018 systematic review reported higher positive predictive values of unexpected WL than we found here.3 This could be accounted for by the considerable heterogeneity between studies included in that review. For example, sensitivity was higher in studies at risk of recall bias. Positive predictive values also varied by the method of data collection and were higher in case-control studies than in cohort studies reporting on the same tumour site. A recent clinical review reported 17 symptoms, signs, and test results, which in combination with unexpected WL had a positive predictive value for cancer of more than 3%.3 These estimates were taken from case-control studies using primary care records that included clinical features occurring in the 1-2 years before the diagnosis of a specific cancer. We studied a much shorter interval around the presentation with unexpected WL, included all cancer sites, and had a study size sufficient to allow separate estimates to be produced for each age group and by smoking status and sex. We found evidence for clinical features and abnormal blood test results that were predictive of cancer in combination with unexpected WL not previously reported: abdominal pain, appetite loss, non-cardiac chest pain, chest signs, dyspepsia, raised alkaline phosphatase levels,
low albumin levels, and a raised white cell count. In addition, our study confirms the importance of jaundice, lymphadenopathy, haemoptysis, dysphagia, thrombocytosis, and anaemia, but the implication of these features co-occurring with unexpected WL differ from when they occur alone.

Conclusions and policy implications
The risk of cancer in patients presenting with unexpected WL alone and who have not smoked, with the exception of men older than 80 years, is below the threshold for referral for intensive cancer investigation set by NICE. However, in combination with the clinical features shown in table 4, the risk of cancer increases such that referral for invasive investigation becomes justified. In the absence of these features, these results might suggest that doctors arrange simple routine blood tests, in particular for a full blood count, liver function, erythrocyte sedimentation rate, C reactive protein, and calcium (figs 3 and 4). Almost any abnormal test result increases the risk of cancer sufficiently to trigger invasive testing. A higher or lower threshold of cancer risk could be chosen to trigger cancer investigation by primary care clinicians practising outside the United Kingdom. The positive predictive values presented will allow clinicians worldwide to use whichever threshold applies locally. However, as negative likelihood ratios were never lower than 0.2, and while normal blood test results might reassure patients, clinicians should be aware that in isolation a normal blood test result does not reduce the probability of cancer downward enough to rule-out the disease in patients with unexpected WL.

A pro-inflammatory state underpins cancer cachexia, and prognostic scores composed of markers of the systemic inflammatory response are used in patients with cachexia to predict survival and response to treatment in secondary care. A potential avenue for future research is to investigate the utility of inflammatory marker scores and combinations of negative test results in selecting who should (and who should not) undergo invasive testing for cancer.

These findings might also have implications for cancer referral pathways. For example, NICE guidelines suggest that patients with unexpected WL and abdominal pain should be investigated for colorectal cancer (supplementary appendix 4). In this study, more than 10 additional cancers presented in this way that would be missed by colonoscopy (fig 3). Likewise, some non-alarm symptoms, such as loss of appetite and non-cardiac chest pain, indicated a probability of cancer that was above the threshold for invasive investigation in the presence of unexpected WL. Throughout Denmark and in some experimental centres in the UK, multidisciplinary diagnostic centres (MDCs) operate that rapidly investigate non-specific symptoms across a broad range of cancer sites. Rapid diagnostic centres are being commissioned throughout the English National Health Service based on the MDC model.

Lastly, women presenting with unexpected WL were at markedly lower risk of having cancer than men with unexpected WL. Different, but interrelated, mechanisms might underpin this finding. Firstly, women might be more likely to visit their doctor to discuss their weight: they are also more likely to have a weight measurement recorded in UK primary care. Secondly, women may be more likely to report earlier symptoms of cancer to prompt investigation before weight loss occurs. Thirdly, men may delay presentation until weight loss is noticeable. This study could not examine these possibilities. Nevertheless, routine weight measurement in primary care could lead to the earlier detection of weight change.

Conclusion
Unexpected WL alone in people who do not smoke is unlikely to be due to cancer and immediate referral for invasive testing might not be justified. In male ever-smokers aged 50 years or older, onward referral might be justified without additional clinical features recorded in the three months before and up to one month after the index date. Some additional clinical features recorded in this time window increase the risk of cancer substantially over the 3% threshold, justifying further investigation. Clinical features thought to be specific to an individual cancer site are markers of several different types of cancer when they co-occur with unexpected WL, which support new, broader investigative approaches for patients with unexpected WL.

We thank David Mant for his expertise and invaluable guidance when developing the fellowship proposal underpinning this research.

Contributors: BDN (principal investigator) conceived and oversaw the study, wrote the protocol, developed the code lists for use in the study, conducted the data management, conducted the statistical analysis with guidance from CK, and interpreted the statistical analysis. SIP collected the CPRD code lists for the features of cancer. BDN wrote the first draft of the manuscript. All members of the team were involved in the drafting and commenting on further revisions of the manuscript. All authors read and approved the final manuscript. BDN is the guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: This study received no specific funding. BDN was supported by National Institute for Health Research (NIHR) doctoral research fellowship (DRF-2015-08-18). PA is an NIHR senior investigator and is funded by NIHR Oxford Biomedical Research Centre (BRC) and Applied Research Collaboration (ARC). WH is co-principal investigator of the multi-institutional CanTest Research Collaborative funded by a Cancer Research UK Population Research Catalyst award (C8640/A233585). FDHR acknowledges part funding from the National Institute for Health Research (NIHR) School for Primary Care Research, the NIHR Collaboration for Leadership in Health Research and Care (CLAHRC) Oxford, the NIHR Oxford BRC and ARC, and the NIHR Oxford Medtech and In-Vitro Diagnostics Co-operative (MIC). While collating the code lists used in this study, SIP was funded by Cancer Research UK funding (C56843/A21550). The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: The protocol was approved by the Independent Scientific Advisory Committee (ISAC) of the Medicines and Healthcare products Regulatory Agency (MHRA) (ISAC protocol No 16_164A2A). Ethical approval for observational research using the CPRD with
Supplementary information: appendices 1-4