
Downloaded from: http://researchonline.lshtm.ac.uk/1883888/
Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults

Krishnan Bhaskaran, Ian Douglas, Harriet Forbes, Isabel dos-Santos-Silva, David A Leon, Liam Smeeth

Summary

Background High body-mass index (BMI) predisposes to several site-specific cancers, but a large-scale systematic and detailed characterisation of patterns of risk across all common cancers adjusted for potential confounders has not previously been undertaken. We aimed to investigate the links between BMI and the most common site-specific cancers.

Methods With primary care data from individuals in the Clinical Practice Research Datalink with BMI data, we fitted Cox models to investigate associations between BMI and 22 of the most common cancers, adjusting for potential confounders. We fitted non-linear (spline) models; investigated effect modification by sex, smoking, and age; and calculated population effects.

Findings 5·24 million individuals were included; 166,955 developed cancers of interest. BMI was associated with 17 of 22 cancers, but effects varied substantially by site. Each 5 kg/m² increase in BMI was roughly linearly associated with cancers of the uterus (hazard ratio [HR] 1.62, 99% CI 1.56–1.69; p<0.0001), gallbladder (1.31, 1.26–1.36; p<0.0001), and leukaemia (1.09, 1.05–1.13; p<0.0001). BMI was positively associated with liver (1.19, 1.12–1.27), colon (1.10, 1.07–1.13), ovarian (1.09, 1.04–1.14), and postmenopausal breast cancers (1.05, 1.03–1.07) overall (all p<0.0001), but these effects varied by underlying BMI or individual-level characteristics. We estimated inverse associations with prostate and premenopausal breast cancer risk, both overall (prostate 0.98, 0.95–1.00; premenopausal breast cancer 0.89, 0.86–0.92) and in never-smokers (prostate 0.96, 0.93–0.99; premenopausal breast cancer 0.89, 0.85–0.94). By contrast, for lung and oral cavity cancer, we observed no association in never smokers (lung 0.99, 0.93–1.05; oral cavity 1.07, 0.91–1.26); inverse associations overall were driven by current smokers and ex-smokers, probably because of residual confounding by smoking amount. Assuming causality, 41% of uterine and 10% or more of gallbladder, kidney, liver, and colon cancers could be attributable to excess weight. We estimated that a 1 kg/m² population-wide increase in BMI would result in 3790 additional annual UK patients developing one of the ten cancers positively associated with BMI.

Interpretation BMI is associated with cancer risk, with substantial population-level effects. The heterogeneity in the effects suggests that different mechanisms are associated with different cancer sites and different patient subgroups.

Funding National Institute for Health Research, Wellcome Trust, and Medical Research Council.

Copyright © Bhaskaran et al. Open Access article distributed under the terms of CC BY.

Introduction

Understanding of the effects of obesity on major health outcomes has never been more urgent, given the rapid rise in obesity worldwide in recent years. Research has suggested that body-mass index (BMI) is an important predictor of cancer risk; a Norwegian cohort study reported associations with several cancer sites, including the thyroid and ovaries; and the UK Million Women Study showed associations between BMI and ten of 17 sites investigated. Two large reviews brought these and many smaller studies together. In a meta-analysis of 221 datasets, strong associations were recorded between BMI and cancers of the oesophagus, thyroid, colon, kidneys, endometrium, and gallbladder, and weaker associations were shown for several other sites. Increased BMI was negatively associated with lung cancer.

However, there are important limitations to the evidence base: individual studies have often had insufficient power, and potential confounders have been inconsistently measured across studies; information is scarce about the role of factors such as age and smoking status that could modify BMI–cancer associations; many studies have used self-reported BMI data, which probably underestimate true BMI; and non-linear associations between BMI and cancer risk have seldom been investigated.

Our aim was to investigate the links between BMI and the most common site-specific cancers using BMI and outcome data from routinely collected UK primary care records. We had the following objectives: to estimate BMI associations with a wide range of cancers with higher precision than has previously been possible; to examine possible non-linear associations between BMI
and cancer; to systematically investigate effect modification by important individual-level factors; and to calculate absolute measures of effect and thus quantify the public health importance of estimated BMI–cancer associations.

**Methods**

**Study design and setting**

We undertook a cohort study with prospectively collected data from the UK Clinical Practice Research Datalink (CPRD), which contains computerised primary care records from general practitioners who use the Vision IT system and have agreed at the practice level to participate (covering about 9% of the UK population). CPRD captures diagnoses, prescriptions, and tests from primary care, and referrals to specialists, hospital admissions, and diagnoses made in secondary care, which are typically reported back to the general practitioners. CPRD has high validity for a range of diagnoses. General practitioners record lifestyle (eg, smoking status, alcohol use) and anthropometric measurements (eg, height, weight); these measurements could be recorded at patient registration, opportunistically during care, or as deemed clinically relevant by the general practitioners. Data collection began in 1987, and we used all data to July, 2012.

**Participants, exposures, and outcomes**

We included all people in CPRD aged 16 years or older with BMI data and subsequent eligible follow-up time available. BMI was calculated directly from weight and height records (weight/height²). We have published details on the processing, cleaning, and representativeness of CPRD BMI data. Exposure was assigned as the earliest BMI recorded during research-standard CPRD follow-up (ie, follow-up meeting CPRD’s data quality criteria). However, to maximise the available follow-up time in individuals without a BMI recorded at the beginning of research-standard follow-up, we used the most recent previous BMI (if available) and time-updated it when the first research-standard BMI record became available (appendix p 6). These older BMIs were dropped in a sensitivity analysis. Other than this specific situation, we did not time-update BMI during follow-up.

Study entry began 12 months after registration because cancers recorded soon after registration could reflect pre-existing or historical disease. Additionally, we assigned BMI records as exposure only 12 months after their recording, to guard against reverse causality (ie, BMI being affected by undiagnosed cancer); this period was extended to 3 years in a sensitivity analysis. Individuals with any record of cancer before study entry were excluded, as were those with data inconsistencies in important dates (date of birth, start and end of follow-up).

To identify outcomes, CPRD clinical records were searched for codes showing malignant disease (appendix p 1). Our outcomes were the 21 most common cancers in the UK (covering 90% of all cancers annually)—namely female breast, prostate, colon, rectum, lung, malignant melanoma, bladder, stomach, oesophageal, non-Hodgkin lymphoma, leukaemia, ovary, pancreas, multiple myeloma, uterus body, brain and central nervous system, liver, kidney, cervix, oral cavity, and thyroid; we included a 22nd cancer type (gallbladder) because of evidence suggesting a link with BMI. More than 90% of nationally registered cancers can be identified in CPRD records, which suggests that it has high sensitivity.

---

![Figure 1: Flow diagram showing the creation of the main dataset, reasons for exclusions, and assignment of body mass index (BMI) at study entry](image-url)
Analyses of female breast cancer were stratified a priori by menopausal status on the basis of previous evidence of different BMI effects. Statistical analysis

People were followed-up from study entry until the earliest of: first cancer diagnosis (any site), death, transfer out of CPRD, or last data collection date for the practice.

We censored 30 days after the first record of hysterectomy for uterine and cervical cancer, and after bilateral oophorectomy for ovarian cancers (we allowed 30 days to capture cancers related to, or detected at, the procedure).

To relate BMI to risk of each cancer, we fitted Cox models with attained age as the underlying timescale.
Articles

Cancer type (ICD10) and number of cases

<table>
<thead>
<tr>
<th>Cancer type (ICD10)</th>
<th>HR (99% CI)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral cavity (C00–06) (7376)</td>
<td>0.82 (0.74–0.89)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Oesophageal (C15) (5213)</td>
<td>1.03 (0.99–1.08)</td>
<td>0.056</td>
</tr>
<tr>
<td>Stomach (C16) (3337)</td>
<td>1.16 (1.09–1.24)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Colon (C18) (31465)</td>
<td>1.03 (0.98–1.09)</td>
<td>0.16</td>
</tr>
<tr>
<td>Rectum (C20) (6123)</td>
<td>1.08 (1.00–1.18)</td>
<td>0.013</td>
</tr>
<tr>
<td>Liver (C22) (1859)</td>
<td>1.10 (1.07–1.13)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Gallbladder (C23) (393)</td>
<td>1.11 (1.07–1.15)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Pancreas (C35) (3951)</td>
<td>1.03 (0.99–1.10)</td>
<td>0.027</td>
</tr>
<tr>
<td>Lung (C34) (19339)</td>
<td>0.82 (0.80–0.85)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Malignant melanoma (C43) (8595)</td>
<td>0.89 (0.86–0.92)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Breast—premenopausal (C50) (5298)</td>
<td>1.05 (1.03–1.07)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Breast—postmenopausal (C51) (28490)</td>
<td>1.05 (1.03–1.08)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Cervix (C53) (1389)</td>
<td>1.10 (1.07–1.13)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Uterus (C54–55) (2758)</td>
<td>1.63 (1.56–1.69)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Ovaries (C56) (3664)</td>
<td>1.09 (1.04–1.14)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Prostate (C61) (24901)</td>
<td>0.98 (0.95–1.00)</td>
<td>0.0042</td>
</tr>
<tr>
<td>Kidney (C64) (1906)</td>
<td>1.25 (1.21–1.29)</td>
<td>0.00036</td>
</tr>
<tr>
<td>Bladder (C67) (7976)</td>
<td>1.03 (0.99–1.06)</td>
<td>0.062</td>
</tr>
<tr>
<td>Brain and CNS (C71–77) (2974)</td>
<td>1.04 (0.99–1.09)</td>
<td>0.053</td>
</tr>
<tr>
<td>Thyroid (C73) (94)</td>
<td>1.02 (0.94–1.10)</td>
<td>0.56</td>
</tr>
<tr>
<td>Non-Hodgkin lymphoma (C82–85) (6946)</td>
<td>1.03 (0.99–1.06)</td>
<td>0.059</td>
</tr>
<tr>
<td>Multiple myeloma (C90) (2969)</td>
<td>1.00 (0.95–1.05)</td>
<td>0.06</td>
</tr>
<tr>
<td>Leukaemia (C91–95) (5833)</td>
<td>1.03 (0.98–1.07)</td>
<td>0.15</td>
</tr>
<tr>
<td>Oesophagus (C15) (5210)</td>
<td>1.05 (0.98–1.11)</td>
<td>0.40</td>
</tr>
<tr>
<td>Stomach (C16) (3337)</td>
<td>1.05 (0.98–1.10)</td>
<td>0.16</td>
</tr>
<tr>
<td>Colon (C18) (31465)</td>
<td>1.08 (1.00–1.18)</td>
<td>0.013</td>
</tr>
<tr>
<td>Rectum (C20) (6123)</td>
<td>1.04 (1.00–1.08)</td>
<td>0.017</td>
</tr>
<tr>
<td>Liver (C22) (1859)</td>
<td>1.19 (1.12–1.27)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Gallbladder (C23) (393)</td>
<td>1.31 (1.25–1.38)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Pancreas (C35) (3951)</td>
<td>1.11 (1.07–1.15)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Lung (C34) (19339)</td>
<td>0.82 (0.80–0.85)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Malignant melanoma (C43) (8595)</td>
<td>0.99 (0.96–1.02)</td>
<td>0.39</td>
</tr>
<tr>
<td>Breast—premenopausal (C50) (5298)</td>
<td>0.96 (0.92–1.00)</td>
<td>0.013</td>
</tr>
<tr>
<td>Breast—postmenopausal (C51) (28490)</td>
<td>0.89 (0.85–0.94)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

For cancers positively associated with BMI, population attributable risks were calculated by fitting a three-category BMI variable (BMI <18·5, 18·5–25, ≥25 kg/m²), and combining estimated hazard ratios (HRs) for overweight and obesity with published national overweight and obesity prevalence (Health Survey for England 2010); we accounted for a potentially different overweight and obesity prevalence (Health Survey for England 2010); and stratified by sex. The assumed causal framework implied by our choice of covariate adjustments is shown in the directed acyclic graph15 (appendix p 7). Because of concerns about residual confounding by smoking we repeated this analysis (post-hoc) restricting to never smokers.

Then, a restricted cubic spline basis for BMI was used to investigate possible non-linearity in each BMI–cancer association (appendix p 1). Effect modification was assessed by introducing interaction terms (one at a time) between the BMI spline basis variables and sex, smoking status, menopausal status (time-updated, postmenopausal defined as aged 50 years or more, which is the approximate average age at menopause in the UK, or ever on hormone replacement therapy), and present age (time-updated, 16–49, 50–64, 65–74, ≥75 years, categories chosen to divide total cancer events into approximate quartiles). If (post-hoc) there was a threshold effect, it was estimated and summarised by plotting piecewise linear models across all possible threshold values and using the one giving the highest likelihood.

For cancers positively associated with BMI, population attributable risks were calculated by fitting a three-category BMI variable (BMI <18·5, 18·5–25, ≥25 kg/m²), and combining estimated hazard ratios (HRs) for overweight and obesity with published national overweight and obesity prevalence (Health Survey for England 2010); we accounted for a potentially different cancer risk in underweight individuals when calculating population attributable risks (appendix p 2).16 Projected

Figure 2: Forest plot of hazard ratios (HR) for each cancer per 5 kg/m² increase in body-mass index (BMI), from models with BMI fitted as a linear effect

Number of incident cancer cases in never smokers only were: oral cavity (302); oesophagus (1858); stomach (1320); colon (6152); rectum (2622); liver (1699); gallbladder (132); pancreas (1525); lung (2674); malignant melanoma (4477); breast—premenopausal (3109); breast—postmenopausal (14833); cervix (5535); uterus (1555); ovaries (1864); prostate (10634); kidney (776); bladder (2687); brain and central nervous system (17359); thyroid (478); non-Hodgkin lymphoma (2122); multiple myeloma (1441); and leukaemia (1685). HRs estimated using a separate model for each cancer with linear BMI term, adjusted for age, diabetes status, smoking, alcohol use, socioeconomic status, calendar year, and stratified by sex; p values from Wald tests on the linear BMI term in each adjusted model.
increases in the number of cancers were estimated under a scenario of a population-wide 1 kg/m² BMI increase as follows: we first replicated our non-linear Cox models with Poisson modelling with additional adjustment for time-updated age, to allow direct prediction of event numbers; we then randomly sampled (with replacement) from the main study population a cohort with the same age and sex distribution as the UK population; we then increased all BMIs by 1 kg/m² and predicted from our models the extra number of cancer events; and finally the percentage increase was applied to present UK cancer incidences to obtain the projected number of extra cancers per year. CIs were estimated by bootstrapping.

We excluded people with missing smoking (49206/5.24 million [0.9%]) and alcohol status (394196/5.24 million [7.5%]). Since 22 cancer outcomes were considered, all CIs are presented at the 99% level.

Model checking and sensitivity analyses

The analysis of effect modification by present age implicitly checks for non-proportional hazards for the BMI variable; we checked for non-proportional hazards in other variables by testing for a zero slope in the scaled Schoenfeld residuals over time.2 In sensitivity analyses, we reinstated the 12 month follow-up after a BMI recording into the analysis to check the effect of this exclusion criterion; extended the exclusion period after a BMI record to 3 years; restricted to patients who had a BMI record soon (<12 months) after registration because these BMI measurements were probably administratively rather than clinically motivated; dropped BMIs recorded before the start of research standard follow-up; used hospital episodes and cause of death data to detect cancers that might have been missed in CPRD; adjusted for general practitioner contact in the first 12 months of CPRD follow-up (as a binary variable), to account for potential selection (collider stratification) bias because of restricting to those with BMI measured (appendix p 8); adjusted for detected non-proportional hazards by adding interactions with present age; and restricted the analysis to recent calendar periods in which BMI completeness in CPRD was higher so any selection biases due to missing BMI would be expected to have less effect. Finally, post-hoc, we re-estimated the association between BMI and postmenopausal breast and ovarian cancers, censoring at first hormone replacement therapy use, in case such treatment masked or diluted BMI–cancer associations. We used hospital episodes and cause of death data to detect cancers that might have been missed in CPRD; adjusted for general practitioner contact in the first 12 months of CPRD follow-up (as a binary variable), to account for potential selection (collider stratification) bias because of restricting to those with BMI measured (appendix p 8); adjusted for detected non-proportional hazards by adding interactions with present age; and restricted the analysis to recent calendar periods in which BMI completeness in CPRD was higher so any selection biases due to missing BMI would be expected to have less effect. Finally, post-hoc, we re-estimated the association between BMI and postmenopausal breast and ovarian cancers, censoring at first hormone replacement therapy use, in case such treatment masked or diluted BMI–cancer associations. We repeated the analysis of liver cancer with multiple imputations to account for missing alcohol data (five imputations, multinomial logistic imputation model including all terms from our substantive model).

The study was approved by the London School of Hygiene and Tropical Medicine Ethics Committee. The prespecified scientific protocol (appendix pp 16–25) was

---

Figure 3: Association between body-mass index (BMI) and specific cancers, allowing for non-linear effects, with 99% CIs

The reference BMI for these plots (with HR fixed as 1·0) was 22 kg/m². Separate models were fitted for each cancer type, each with a restricted cubic spline for BMI (knots placed at equal percentiles of BMI), adjusted for age, calendar year, diabetes status, alcohol use, smoking (all at time of BMI recording), socioeconomic status (index of multiple deprivation), and stratified by sex. HR=hazard ratio.

---

For the Health Survey for England 2010 see http://www.hscic.gov.uk/pubs/hse10trends Article
produced by adding interaction terms with the BMI spline basis. For estimated effect modification by sex, diabetes status, smoking, alcohol use, socioeconomic status (index of multiple deprivation). Stratified curves were estimated from models with BMI fitted as a spline, adjusted for age, calendar year, optimal knots or thresholds). HR=hazard ratio. 5 kg/m² derived from best fitting piecewise linear or linear model (with Akaike information criterion used to select smoking, menopausal status, and present age for all cancer types, see appendix pp 9–12. Estimated HRs per prostate cancers and malignant melanoma, including detected non-linearities and effect modification by sex, smoking, alcohol use, socioeconomic status (index of multiple deprivation). Mean BMI was 25·5 kg/m² (SD 4·9). Compared with the overall CPRD population, those included were more likely to be female (2·86 million/5·24 million [54·6%] vs 5·29 million/10·04 million [52·7%] overall), and to have an earlier year of birth (median 1962 [IQR 1946–1974] vs 1964 [1946–1977] overall). Table 1 and appendix (p 3) describe the characteristics of those included and excluded in the analysis. 201504 of 5·24 million individuals (3·8%) developed any cancer and 166955 (3·2%) developed one of the 22 cancers of interest during follow-up. Appendix (p 4) presents the numbers for each cancer site and relative risks by WHO obesity category adjusted for age and sex. For 13 of 22 cancers, one or more of the overweight and obese categories seemed to be associated with higher risk compared with normal weight, but the size of the effect and patterns across BMI categories varied by site (appendix p 4). Underweight seemed to be associated with increased risk of oral cavity, oesophageal, stomach, and lung cancers, but decreased risk of postmenopausal breast and prostate cancers (appendix p 4).

With BMI included as a linear effect, and adjusting for all potential confounders, each 5 kg/m² increase in BMI was associated with a large increase in risk of cancer of the uterus (HR 1·62, 99% CI 1·56–1·69; figure 2). There were also large increases in risk (HR >1·1 per 5 kg/m²) for cancers of the gallbladder, kidney, and liver, and small increases in risk for colon, cervical, thyroid, ovarian, and postmenopausal breast cancers and leukaemia (figure 2). Net inverse associations were estimated for oral cavity, lung, premenopausal breast, and prostate cancers, but restricting to never smokers, the inverse association with lung and oral cavity cancers disappeared (figure 2). There was a positive association between BMI and oesophageal and stomach cancers in never smokers; other effect estimates were similar in never smokers and the full population (figure 2).

Figure 3 shows the estimated shape of each BMI–cancer association, allowing for non-linearity. These non-linear models were refitted with interaction terms to assess possible modification by sex, smoking status, menopausal status, and present age (appendix pp 9–12). These analyses showed overall evidence of association between BMI and...
17 of 22 cancers studied (all except rectum, bladder, brain and CNS, non-Hodgkin lymphoma, multiple myeloma), and evidence of non-linearity or effect modification for ten cancers (figure 3; appendix pp 9–12). Associations between BMI and both colon and liver cancer were more marked in men than women: in men, higher BMI was associated with substantially increased risk but only if BMI was above 22 kg/m², whereas in women more modest increases in risk were recorded with no evidence against linearity (figure 4). Increases in ovarian cancer risk with BMI were larger in premenopausal than postmenopausal women. As expected, there were differences by menopausal status for breast cancer: BMI was positively associated with both premenopausal and postmenopausal breast cancers at the lower end of the BMI range (<22 kg/m²), whereas above this value, risk of premenopausal breast cancer reduced markedly with increasing BMI (figure 4). A similar peaked pattern was noted for prostate cancer and malignant melanoma in men (with risk peaking at BMI 27 kg/m² for prostate cancer and 24 kg/m² for malignant melanoma; figure 4).

For lung, oral cavity, and stomach cancers, low BMI was associated with increased risk but this risk was driven by current smokers and ex-smokers and was attenuated or disappeared in never smokers (figure 5). Smoking seemed to drive the apparent effect modification by sex and age for these cancers (p for interaction >0.1 in each case when restricted to never smokers; appendix pp 9, 12). For oesophageal cancer, there was evidence of effect modification by sex, smoking, and age (appendix pp 9–12).

With the assumption that the recorded associations were causal, we estimated that 41% of cancers of the uterus and more than 10% of gallbladder, kidney, liver, and colon cancers are attributable to overweight and obesity, and that a 1 kg/m² population-wide increase in BMI would lead to 3790 extra patients developing one of the ten cancers that had a net positive association with BMI (113 928 projected per year in total compared with 110 138 at present; table 2). For the two cancers overall inversely associated with BMI even in never-smokers, the corresponding projected decreases in numbers of new diagnoses per year were modest (prostate: projected change –72, 99% CI –159 to 15; premenopausal breast: –128, –147 to –110).

Our main findings were robust to a range of sensitivity analyses; estimated effect sizes were similar and in all cases, CIs included the point estimate from the main analysis (appendix pp 5, 13–15).

**Discussion**

In the largest single dataset assembled so far for this purpose, we recorded associations between BMI and 17 of 22 cancers studied, but effects varied substantially by cancer type, in both direction and size. Higher BMI was roughly linearly related with increased risk of uterine, gallbladder, kidney, cervical, thyroid cancers, and leukaemia. Overall positive associations were recorded between BMI and liver, colon, ovarian, and postmenopausal breast cancers, but these associations were non-linear (the effect of BMI varying across the BMI range) or modified by individual level factors. BMI had a net inverse association with risk of premenopausal breast and prostate cancers; for lung and oral cavity cancers, an overall inverse association seemed to be driven by smoking and was not observed in never smokers.
Renehan and colleagues’ summarised the evidence about BMI and cancer in their landmark 2008 systematic review and meta-analysis. In our study we have added to this evidence by systematically investigating non-linearity and effect modification by individual factors. These analyses mean that we were able to characterise relationships between BMI and a wide range of individual cancers in greater detail than has previously been possible. To our knowledge, ours is the first large-scale study of its kind to estimate public health effects of BMI–cancer associations across a range of sites.

For the six cancers positively and linearly associated with BMI, our effect sizes were broadly consistent with previous studies (appendix pp 27–30). We did not find strong evidence of non-linearity in the association between BMI and cancer of the uterus, although our estimated curve was not incompatible with a larger effect at higher BMI, as was reported in one previous study.5,20,21

Another explanation is residual confounding by amount of smoking: heavy drinking could be associated with both alcohol use and smoking, and thus appear to mediate the association between BMI and cancer. 

For the other ten cancers, we confirmed previously identified links with BMI, but we identified non-linear relationships or important effect modifications that had not previously been characterised in detail. The modest overall effect of BMI on ovarian cancer risk matched that in previous studies,22 but we noted a much larger effect in premenopausal compared with postmenopausal women; this finding contrasts with those of a meta-analysis which showed no difference when combining six previous studies that stratified by menopausal status, although a weakness was the different definitions of menopausal status used across the constituent studies.22 Nevertheless, given the conflicting evidence, this finding of effect modification needs to be replicated (or otherwise) elsewhere.

BMI has been inversely associated with premenopausal and positively associated with postmenopausal breast cancer,1 which we also noted, but incorporating non-linearity showed a more nuanced picture: risk of premenopausal breast cancer seemed to peak at 22 kg/m² then dropped as BMI increased further. A similar pattern was evident for malignant melanoma and prostate cancer in men, which could explain why previous studies treating BMI as a linear term have estimated only modest or inconsistent associations with these cancers.7 This might reflect a real non-linear biological association, but another explanation is that diagnoses could be delayed or missed in people who are overweight and obese, which is supported by the finding that BMI correlates inversely with age and sex structure of the UK population, with all BMIs increased by 1 kg/m², and applying proportionate increase to the annual number of UK cases (*Cancer Research UK cancer statistics*).99% CIs are from a bootstrapping procedure.

### Table 2: Estimated UK population effects for cancers with evidence of a net positive association with body-mass index (BMI)

<table>
<thead>
<tr>
<th>Cancer Site</th>
<th>New cases per year (UK)*</th>
<th>n (%) cases attributable to overweight and obesity</th>
<th>Projected extra cases per year with a 1 kg/m² population-wide increase in BMI (99% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colon (C18)</td>
<td>26,725</td>
<td>29,70 (11.1%)</td>
<td>559 (519–598)</td>
</tr>
<tr>
<td>Liver (C22)</td>
<td>4,241</td>
<td>651 (15.6%)</td>
<td>145 (135–154)</td>
</tr>
<tr>
<td>Gall bladder (C23)</td>
<td>660</td>
<td>134 (20.3%)</td>
<td>36 (35–37)</td>
</tr>
<tr>
<td>Breast (post-menopausal, C50)</td>
<td>39,812</td>
<td>2035 (5.1%)</td>
<td>1,441 (1,417–1,465)</td>
</tr>
<tr>
<td>Cervix (C53)</td>
<td>2,851</td>
<td>214 (7.5%)</td>
<td>51 (50–53)</td>
</tr>
<tr>
<td>Uterus (C54–55)</td>
<td>8,288</td>
<td>338 (4.0%)</td>
<td>806 (784–829)</td>
</tr>
<tr>
<td>Ovaries (C56)</td>
<td>7,011</td>
<td>512 (7.3%)</td>
<td>125 (118–133)</td>
</tr>
<tr>
<td>Kidney (C64)</td>
<td>9,639</td>
<td>1,597 (16.6%)</td>
<td>428 (414–442)</td>
</tr>
<tr>
<td>Thyroid (C73)</td>
<td>2654</td>
<td>51 (1.9%)</td>
<td>49 (48–51)</td>
</tr>
<tr>
<td>Leukaemia (C91–95)</td>
<td>8,257</td>
<td>522 (6.3%)</td>
<td>150 (138–163)</td>
</tr>
</tbody>
</table>

Attributable cases and percentage based on models for each cancer with a three-category BMI variable (underweight, normal, or overweight and obese), adjusted for age, diabetes status, smoking, alcohol use, socioeconomic status, calendar year, and stratified by sex; attributable cases were calculated separately for men and women and then combined; we assumed the prevalence of overweight including obese to be 65% in men and 58% in women (Health Survey for England 2010*). Further details on our method for these calculations are given in appendix (p 2). Projected extra cases calculated with fitted Poisson models with non-linear BMI splines to predict proportion of extra cases in an artificial population sampled from the original cohort to replicate the age and sex structure of the UK population, with all BMIs increased by 1 kg/m², and applying proportionate increase to the annual number of UK cases (*Cancer Research UK cancer statistics*).99% CIs are from a bootstrapping procedure.
sex hormones, have all been implicated, as have adipokines (signalling proteins secreted by adipose tissue),
but the precise roles of these mechanisms and the interactions between them are not completely understood. The diabetes-associated increase in mortality from certain cancers seemed to persist after adjustment for BMI in one study, suggesting that BMI might be upstream of diabetes and glycaemia as part of one causal chain. Large-scale biobank data linked to long-term outcomes might eventually help to further clarify these multiple potential pathways.

Our study has considerable strengths. We used a large data source to estimate associations with unprecedented precision and power; we applied a consistent methodological approach to examine BMI effects across cancer sites, including systematic investigation of non-linearity and effect modification. The validity of CPRD diagnosis data has been established in general, and for cancer in particular (>90% of CPRD cancers confirmed from other sources, >90% of nationally registered cancers present in CPRD). Although some outcome misclassification was inevitable, the effect on our results was probably small, and an additional sensitivity analysis using hospital and death certificate data to capture cancers that might have been missed in CPRD gave similar results to our main analysis (appendix p 13). CPRD patients broadly represent the wider population and are not restricted on age or other factors, suggesting good generalisability to the UK population and to comparable countries. Our findings were robust in a wide range of sensitivity analyses.

There are important limitations. Individuals without BMI data were not included; the decision to measure BMI in primary care might be related to the patient’s apparent weight or their health status, introducing the potential for selection bias in our HRs and projections of public health effect. We already reported (allowing for differences in age and sex) that those with up-to-date BMI data in CPRD had mean BMI and obesity prevalence close to that expected based on representative population-based survey data, suggesting that those with complete data in CPRD were not substantially different in terms of BMI to the broader population. We were empirically reassured against serious selection bias by the consistency of our results with previous research, including cohort studies with complete BMI information (appendix pp 27–30). In a sensitivity analysis, we used only BMIs recorded within 12 months of general practitioner registration (because they would have probably been recorded for administrative and not health reasons) and obtained similar results (appendix p 13). In a further sensitivity analysis, we increasingly restricted to more recent follow-up when BMI completeness was higher (including the period from 2005 onwards during which BMI recording by general practitioners was financially incentivised through the UK Quality and Outcomes Framework), and recorded very little change in our estimates (appendix p 14). To induce selection bias, having a BMI measurement would need to be related to both BMI and later cancer risk; this could be the case if general practitioner attenders tend to have different cancer risk (appendix p 8), or indeed a different probability of cancer being diagnosed (eg, through better access to screening services), but a sensitivity analysis adjusting for this had no material effect on the conclusions. Our use of complete case analysis to assess missing BMI data would be valid providing that the probability of having BMI recorded is independent of the outcome conditional on covariates; more intuitively, the BMI–cancer relation should be similar between included and excluded individuals (after adjustment for our covariate set). We had no a-priori reason to doubt this assumption, and we are reassured that our main conclusions are unlikely to have been seriously affected by bias due to missing BMI measurements.
A second important limitation is the potential for residual confounding. Despite adjustment for smoking status (which itself could have been misclassified to some extent), there was probably confounding by amount smoked, although restricting to never smokers gave similar estimates to the adjusted analysis in the full population for all but four strongly smoking-related cancers (lung, oral cavity, stomach, oesophageal). Alcohol status was likely to be self-reported to the general practitioner (prone to misclassification), and information was scarce about the amount of drinking. We had no information about physical activity (although its effects on cancer risk might be in part mediated through BMI itself), on female reproductive factors such as parity and age at first birth, or on potentially important infections (eg, hepatitis B and C for liver cancer). Our outcome data were insufficiently detailed to investigate potentially important differences between cancer subtypes, in particular for oesophageal cancer (adenocarcinoma vs squamous cell), breast cancer (by receptor status), non-Hodgkin lymphoma (follicular vs diffuse large B-cell), and colon cancer (microsatellite stable vs unstable, or distal vs proximal). We examined only BMI associations with first cancer and censored follow-up at this point, so our study does not provide information about associations between BMI and risk of second or subsequent cancers; furthermore, this censoring could introduce bias if certain cancers tend to be affected by BMI sooner than others, but since 92% of individuals with a cancer diagnosis had only one such diagnosis, censoring is unlikely to have had any substantial effect. Finally, we had no data for important measures of adiposity other than BMI.

There is no systematic analysis of non-linearity and effect modification across a wide range of cancer sites with which to directly compare our findings (panel); this emphasises the novelty of our study but means that new findings from such analyses will need to be reproduced by others before they can be considered definitive.

Assuming the relationships to be causal, many cancers are attributable to overweight and obesity. Even within normal BMI ranges, higher BMI was associated with increased risk of some cancers, accentuating the public health implications in view of the overall increase in population BMI distributions in several countries. We estimated that a 1 kg/m² population increase in mean BMI would lead to 3790 additional patients in the UK each year developing one of the ten cancers that were overall positively associated with BMI. To put this in context, based on data from 2003 to 2010, mean BMI in England has been increasing at a rate equivalent to a 1 kg/m² increase taking around 12 years. Our data strengthen the rationale to assess and implement strategies aimed at stopping these trends and mitigating their public health effects.

Contributors
All authors were involved in the study design. KB did the statistical analysis. HF did the systematic literature search for the Research in Context section. KB wrote the first draft. All authors contributed to further drafts and approved the final manuscript.

Declaration of interests
We declare no competing interests.

Acknowledgments
This report is independent research arising from a postdoctoral fellowship (for KB) supported by the National Institute for Health Research. LS is supported by a Wellcome Trust senior research fellowship in clinical science. ID is supported by an MRC methodology research fellowship. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health.

References


