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 linear then non-linear (spline) models; investigated effect modification by sex, menopausal status, 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, 1·12–1·52; p<0·0001), kidney (1·25, 1·17–1·33; p<0·0001), cervix (1·10, 1·03–1·17; p=0·00035), thyroid (1·09, 1·00–1·19; p=0·0088), 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.

Lancet 2014; 384: 755–65

Published Online
August 14, 2014
http://dx.doi.org/10.1016/S0140-6736(14)60892-8

See Comment page 727

Department of Non-Communicable Diseases Epidemiology , London School of Hygiene and Tropical Medicine, London, UK (K Bhaskaran PhD, I Douglas PhD, H Forbes MSc, Prof I dos-Santos-Silva PhD, Prof D A Leon PhD, Prof L Smeeth PhD); **and Farr Institute of Health Informatics Research , London, UK** (Prof L Smeeth)

Correspondence to:
Dr Krishnan Bhaskaran, Department of Non-Communicable Diseases Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK
krishnan.bhaskaran@lshtm.ac.uk

Introduction

Understanding of the effects of adiposity on major health outcomes has never been more urgent, given the rapid rise in obesity worldwide in recent years.[1] Research has suggested that body-mass index (BMI) is an important predictor of cancer risk:[2] a Norwegian cohort study reported associations with several cancer sites, including the thyroid[3] and ovaries;[4] and the UK Million Women Study showed associations between BMI and ten of 17 sites investigated.[5] Two large reviews brought these and many smaller studies together.[6,7] 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.[7] 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;[8] 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

See Online for appendix

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.[9] 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.

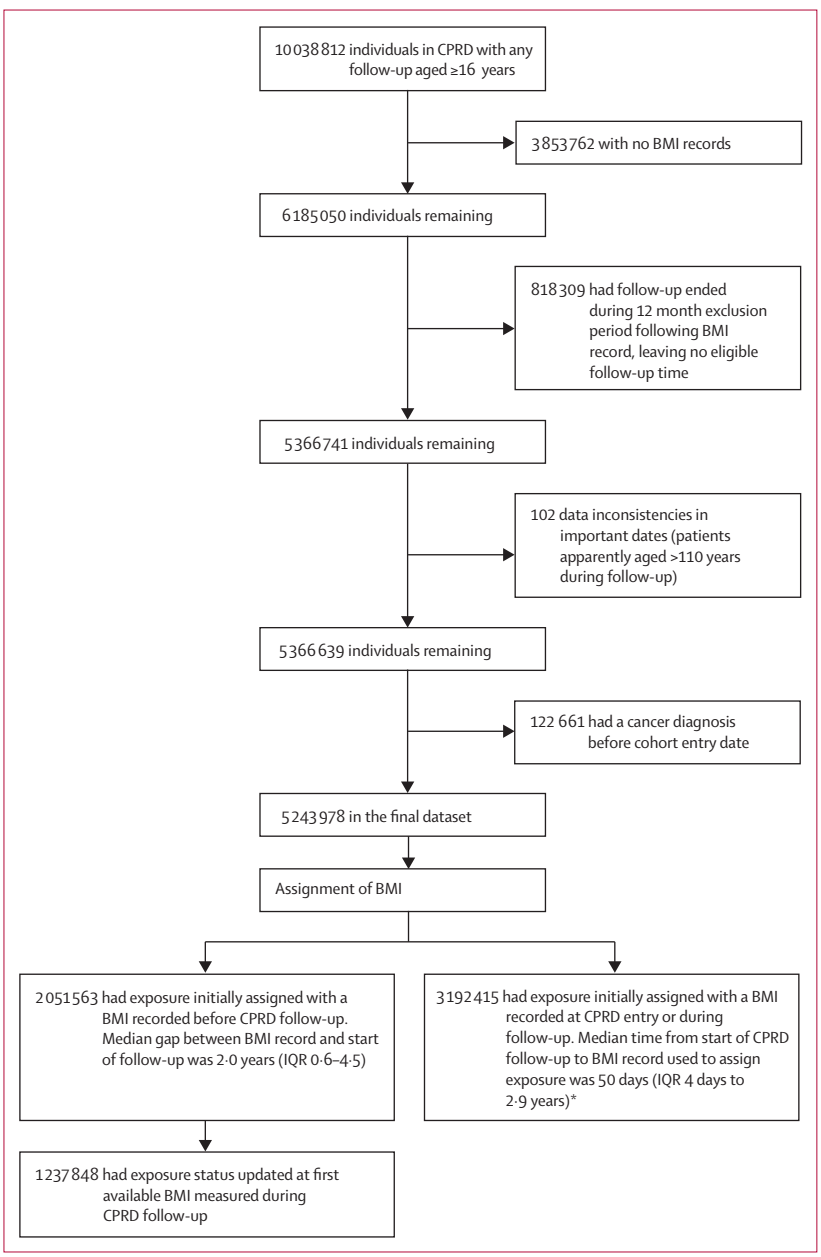

Figure 1: **Flow diagram showing the creation of the main dataset, reasons for exclusions, and assignment of body-mass index (BMI) at study entry**
CPRD= Clinical Practice Research Datalink. *When the first available BMI was after start of CPRD follow-up, the patient was late-entered into the risk set.

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.[10] 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;[11] we included a 22nd cancer type (gallbladder) because of evidence suggesting a link with BMI.[7] More than 90% of nationally registered cancers can be identified in CPRD records, which suggests that it has high sensitivity.[12]

	Underweight (BMI <18·5 kg/m²)	Normal BMI (18·5–25 kg/m²)	Overweight and obese (BMI ≥25 kg/m²)	Overall
N	165 530	2 571 573	2 506 875	5 243 978
Person-years from BMI date to end of follow-up				
Mean (SD)	6·2 (4·8)	7·4 (5·4)	7·6 (5·2)	7·5 (5·3)
Median (IQR)	4·7 (2·4–8·9)	5·9 (2·9–11·1)	6·4 (3·2–11·1)	6·0 (3·0–11·0)
Range	1·0–24·5	1·0–24·5	1·0–24·6	1·0–24·6
Total person-years included* (millions)	0·869	16·451	16·521	33·841
Age (years)				
Median (IQR)	25·5 (19·5–39·4)	33·4 (24·6–48·4)	43·6 (31·5–57·6)	37·9 (27·0–53·7)
Sex				
Male	49 107 (29·7)	1 038 167 (40·4)	1 292 046 (51·5)	2 379 320 (45·4)
Female	116 423 (70·3)	1 533 406 (59·6)	1 214 829 (48·5)	2 864 658 (54·6)
Smoking status				
Never smoker	75 464 (45·6)	1 236 201 (48·1)	1 167 989 (46·6)	2 479 654 (47·3)
Current smoker	69 580 (42·0)	961 893 (37·4)	830 733 (33·1)	1 862 206 (35·5)
Ex-smoker	17 364 (10·5)	348 455 (13·6)	487 093 (19·4)	852 912 (16·3)
Missing	3122 (1·9)	25 024 (1·0)	21 060 (0·8)	49 206 (0·9)
Alcohol use†				
Non-drinker	33 778 (20·4)	315 463 (12·3)	303 750 (12·1)	652 991 (12·5)
Current drinker (light)	80 323 (48·5)	1 519 351 (59·1)	1 401 081 (55·9)	3 000 755 (57·2)
Current drinker (moderate)	8133 (4·9)	243 569 (9·5)	282 474 (11·3)	534 176 (10·2)
Current drinker (heavy)	4499 (2·7)	64 177 (2·5)	70 285 (2·8)	138 961 (2·6)
Current drinker (amount not known)	6414 (3·9)	96 109 (3·7)	98 992 (3·9)	201 515 (3·8)
Ex-drinker	10 043 (6·1)	136 928 (5·3)	174 413 (7·0)	321 384 (6·1)
Missing	22 340 (13·5)	195 976 (7·6)	175 880 (7·0)	394 196 (7·5)
Previous diabetes	1965 (1·2)	46 330 (1·8)	143 033 (5·7)	191 328 (3·6)
Index of multiple deprivation quintile				
1 (least deprived)	20 526 (12·4)	390 322 (15·2)	331 347 (13·2)	742 195 (14·2)
2	42 175 (25·5)	738 004 (28·7)	697 055 (27·8)	1 477 234 (28·2)
3	46 881 (28·3)	726 831 (28·3)	734 314 (29·3)	1 508 026 (28·8)
4	35 950 (21·7)	487 334 (19·0)	496 929 (19·8)	1 020 213 (19·5)
5 (most deprived)	19 998 (12·1)	229 082 (8·9)	247 230 (9·9)	496 310 (9·5)
Calendar year				
<1989	2139 (1·3)	42 548 (1·7)	36 460 (1·5)	81 147 (1·5)
1990–94	32 163 (19·4)	666 646 (25·9)	549 264 (21·9)	1 248 073 (23·8)
1995–99	34 157 (20·6)	584 901 (22·7)	547 819 (21·9)	1 166 877 (22·3)
2000–04	40 390 (24·4)	572 579 (22·3)	618 579 (24·7)	1 231 548 (23·5)
2005–09	46 310 (28·0)	584 572 (22·7)	621 717 (24·8)	1 252 599 (23·9)
≥2010	10 371 (6·3)	120 327 (4·7)	133 036 (5·3)	263 734 (5·0)

Data are n (%) unless otherwise stated. Time-varying characteristics are at date of first eligible BMI measurement. *After excluding first 12 months after registration, and any time without a mature (≥12 months old) BMI measurement available. †Light, moderate, or heavy current drinking were assigned when either the general practitioner had recorded a Read code with one of these terms, or when the units per day were recorded as 1–2 (light), 3–6 (moderate), 7+ (heavy).

Table 1: **Characteristics of individuals included in the study, overall and by body-mass index (BMI) category**

Analyses of female breast cancer were stratified a priori by menopausal status on the basis of previous evidence of different BMI effects.[5,7] At the first diagnosis of any cancer (including sites not investigated here), patients were censored for other cancer sites, because of difficulty in separating metastatic from second de-novo cancers, and the different risk profile of cancer survivors. The detailed algorithms used to process and derive variables in our analysis are available on request from the corresponding author.

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. We

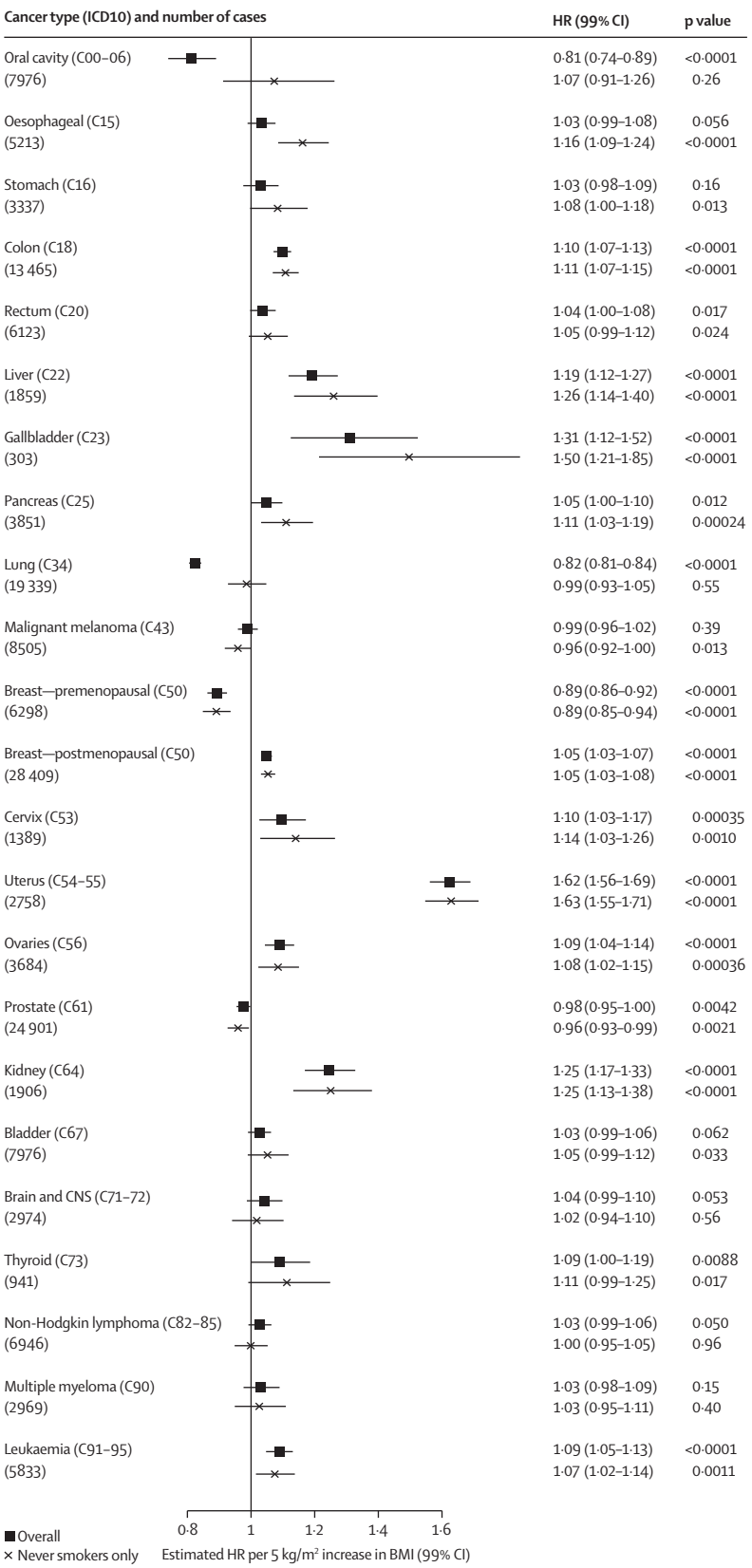

Figure 2: **Forest plot of hazard ratios (HR) for each cancer per 5 kg/m² increase in body-mass index (BMI), from models with BMI fitted as a linear effect**

Number of incident cancer cases in never smokers only were: oral cavity (302); oesophagus (1858); stomach (1320); colon (6115); rectum (2623); liver (699); gallbladder (133); pancreas (1525); lung (2674); malignant melanoma (4477); breast—premenopausal (3109); breast—postmenopausal (14 833); cervix (535); uterus (1555); ovaries (1864); prostate (10 634); kidney (776); bladder (2687); brain and central nervous system (CNS) (1359); thyroid (478); non-Hodgkin lymphoma (3212); multiple myeloma (1441); and leukaemia (2685). 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.

used the same systematic analysis strategy consistently across cancer sites. We initially adjusted for age at BMI record and sex only, and considered BMI in WHO categories.[13] We then fitted fully adjusted models, with BMI as a continuous linear term to estimate the average effect of a 5 kg/m² increase in BMI on cancer risk; we controlled for the following covariates at time of the BMI record(s): age (three-knot restricted cubic spline to allow for non-linearity); smoking status (never smoker, current smoker, ex-smoker); alcohol use (non-drinker, current drinker [light, moderate, heavy, unknown], ex-drinker); previous diabetes diagnosis; index of multiple deprivation (in quintiles, a measure of socioeconomic status);[14] calendar period (<1989, 1990–94, 1995–99, 2000–04, 2005–09, ≥2010); and stratified by sex. The assumed causal framework implied by our choice of covariate adjustments is shown in the directed acyclic graph[15] (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 fitting 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

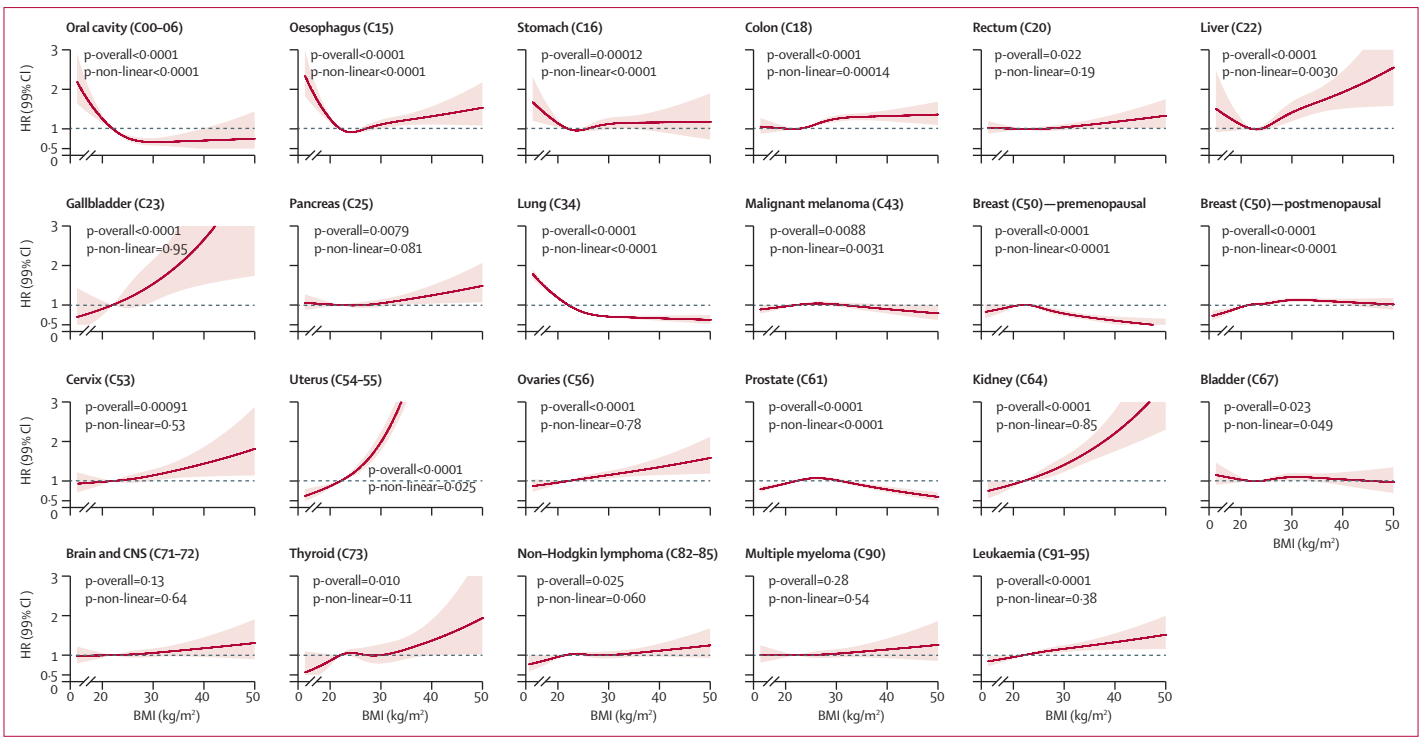

Figure 3: **Association between body-mass index (BMI) and specific cancers, allowing for non-linear effects, with 99% CIs**
The reference BMI for these plots (with HR fixed as 1·0) was 22 kg/m². Separate models were fitted for each cancer type, each with a restricted cubic spline for BMI (knots placed 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.

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 (49 206/5·24 million [0·9%]) and alcohol status (394 196/5·24 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.[17] 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[18] 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

For the **Health Survey for England 2010** see http://www.hscic.gov.uk/pubs/hse10trends

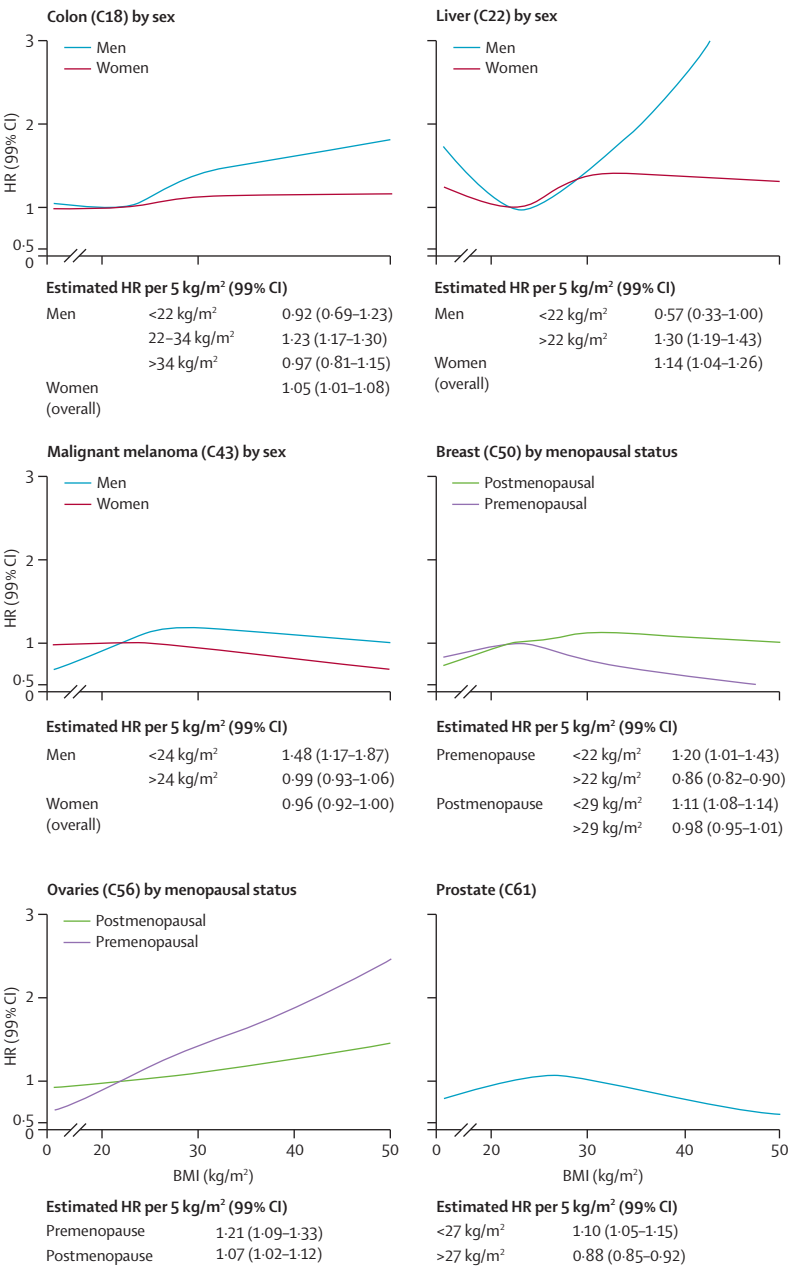

Figure 4: **Modelled associations between body-mass index (BMI) and colon, liver, breast, ovarian, and prostate cancers and malignant melanoma, including detected non-linearities and effect modification**
Curves for each cancer type estimated from models with BMI fitted as a spline, adjusted for age, calendar year, diabetes status, smoking, alcohol use, socioeconomic status (index of multiple deprivation). Stratified curves were produced by adding interaction terms with the BMI spline basis. For estimated effect modification by sex, smoking, menopausal status, and present age for all cancer types, see appendix pp 9–12. Estimated HRs per 5 kg/m² derived from best fitting piecewise linear or linear model (with Akaike information criterion used to select optimal knots or thresholds). HR=hazard ratio.

approved by the Independent Scientific Advisory Committee for MHRA Database Research in July, 2012. Data analysis was done in Stata version 12, on a high-performance computer cluster.

Role of the funding source
The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results
Of 10·04 million individuals aged 16 years or older in CPRD, 6·19 million had any BMI measurement recorded. Of these, 5·24 million (84%) had eligible follow-up time and no previous cancer diagnosis and were included in the study (figure 1); follow-up ended a mean of 7·5 years (SD 5·3) after the first eligible BMI measurement. 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.

201 504 of 5·24 million individuals (3·8%) developed any cancer and 166 955 (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).

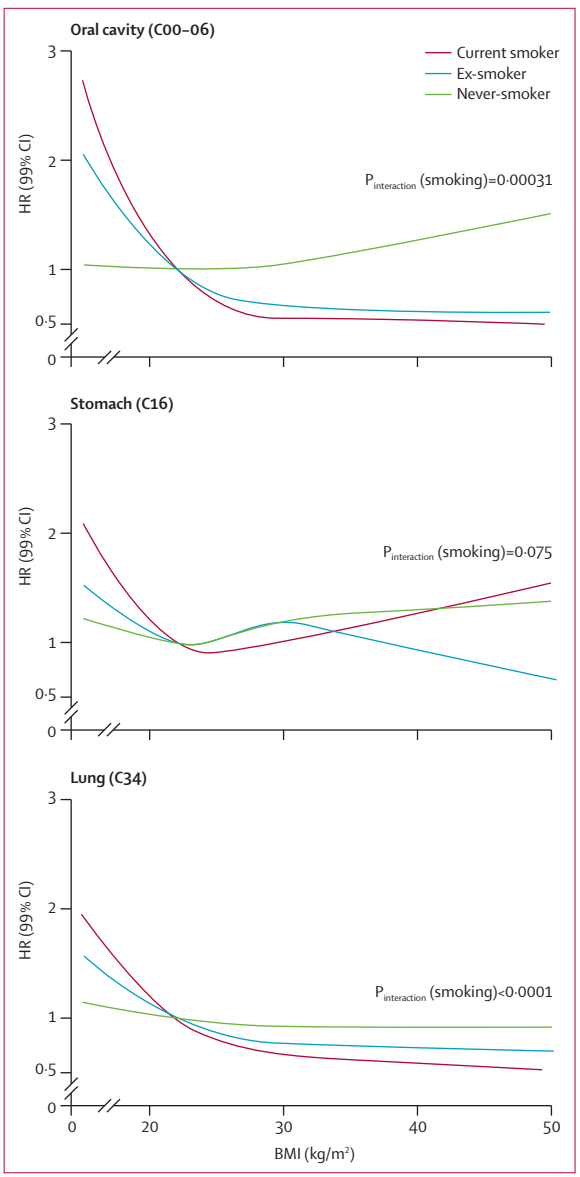

Figure 5: **Associations between body-mass index and oral, stomach, and lung cancers with effect modification by smoking status**
Curves for each cancer type estimated from models with BMI fitted as a spline, interaction terms between smoking status and spline basis, adjusted for age, calendar year, diabetes status, alcohol use, socioeconomic status (index of multiple deprivation), and stratified by sex. p values for effect of BMI in never smokers=0·62 for oral cavity cancer, 0·16 for stomach cancer, and 0·18 for lung cancer. Estimated curves by smoking status for all cancer types are presented in appendix p 10. $P_{interaction}$=p value for interaction. HR=hazard ratio.

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.

	New cases per year (UK)*	n (%) cases attributable to overweight and obesity	Projected extra cases per year with a 1 kg/m² population-wide increase in BMI (99% CI)
Colon (C18)	26 725	2970 (11·1%)	559 (519–598)
Liver (C22)	4241	661 (15·6%)	145 (135–154)
Gall bladder (C23)	660	134 (20·3%)	36 (35–37)
Breast (post-menopausal, C50)	39 812	2035 (5·1%)	1441 (1417–1465)
Cervix (C53)	2851	214 (7·5%)	51 (50–53)
Uterus (C54–55)	8288	3384 (40·8%)	806 (784–829)
Ovaries (C56)	7011	512 (7·3%)	125 (118–133)
Kidney (C64)	9639	1597 (16·6%)	428 (414–442)
Thyroid (C73)	2654	51 (1·9%)	49 (48–51)
Leukaemia (C91–95)	8257	522 (6·3%)	150 (138–163)

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[16]). 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[11]); 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)**

Renehan and colleagues[7] 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 meta-analysis.[19] There was strong evidence of a modest positive association of BMI with cervical cancer, a site which was not included in the meta-analysis by Renehan and colleagues:[7] the estimated 10% increase in risk per 5 kg/m² was consistent with the CIs from the three other studies we identified in our updated systematic review that investigated this site (panel; appendix p 29),[5,20,21] although in only one of these did CIs for the HR exclude 1.[21] We detected no strong evidence of associations between BMI and cancers of the rectum, brain, CNS, and bladder, non-Hodgkin lymphoma, or multiple myeloma, and only weak evidence for cancer of the pancreas; for some of these sites, small associations have been reported previously, but these were consistent with our CIs (appendix pp 27–30).

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,[7,22] 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,[5] 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 localised prostate cancer and positively with advanced disease.[23]

The inverse relationship between BMI and lung cancer concurs with other research,[7] and we noted a similar pattern for oral cavity cancer—the effect was predominantly at the lower end of the BMI scale, and was evident only in current smokers or ex-smokers (as in another recent study of lung cancer[24]). Confounding by amount of smoking is the most likely explanation (low BMI acting as a proxy for heavier smoking) because we had only crudely categorised smoking data. By contrast, high risk of oesophageal cancer in underweight individuals was evident even in never smokers. However, we had no information about oesophageal cancer type; evidence suggests that BMI could have opposite effects on adenocarcinoma versus squamous cell carcinoma,[7] which might have resulted in the U-shaped association overall. Another explanation is residual confounding by amount of drinking: heavy drinking could be associated with both underweight and overweight in different individuals.

Heterogeneity in the effects of BMI suggests that there are different mechanisms or combinations of mechanisms associated with different sites and in different patient subgroups. Several pathways have been proposed; changes in hormone metabolism, particularly with regard to insulin, insulin-like growth factors, and

sex hormones, have all been implicated, as have adipokines (signalling proteins secreted by adipose tissue),[25] 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,[26] 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,[9] and for cancer in particular (>90% of CPRD cancers confirmed from other sources, >90% of nationally registered cancers present in CPRD).[12,27] 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[28] and are not restricted on age or other factors, suggesting good generalisibility 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,[10] 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;[29] 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.

Panel: Research in context

Systematic review

The most relevant summary of the literature on BMI and cancer risk was a 2008 systematic review published in *The Lancet* examining associations between BMI and several cancer sites, and differences in these associations by sex.[7] To update this Review, we searched Medline for reports published in any language since 2008, relating BMI to these cancers, with the same search keywords described by the authors (terms referring to bodyweight combined with terms for each cancer site). We used a similar search strategy to look for research on BMI and bladder, brain and central nervous system, cervical, and oral cavity cancers, which were not in the 2008 review. The search was supplemented with hand searches of reference lists of relevant papers. We included studies of prospectively collected data, in which BMI was treated as a continuous exposure and one or more of the cancers of interest were included as outcomes. We excluded studies in which only BMI categories were considered. We rescaled all reported linear effects to a 5 kg/m² BMI increase. We identified 28 studies to supplement Renehan and colleagues' systematic review (appendix p 26). The main features and findings of these studies are reported by cancer site in appendix (pp 27–30) and compared with our own findings. The published evidence so far suggests associations between BMI and a cancer that differ widely by cancer site in both size and character. We identified very few data for possible non-linearities. Effect modification by sex was examined in a few studies, and most findings showed a larger effect on colon cancer in men than women; information was scarce across the range of cancers on effect modification by smoking, menopausal status, and age. We did not identify any studies that reported BMI associations with oral cavity cancer.

Interpretation

Our data suggest wide-ranging associations between BMI and cancer that vary by cancer site; the linear effects we estimated were consistent with those that have been reported previously (when available). Our study seems to be the first to have systematically investigated non-linearity and the role of individual-level factors (sex, smoking, menopausal status, and age) across a wide range of cancers, allowing us to better understand the relationship between BMI and cancer risk; for ten of 22 cancers we detected important non-linearity or effect modification. The use of a large data source enabled us to include cancers that are not reported in the literature on BMI and cancer risk, and to provide statistically precise effect estimates confirming associations between BMI and liver and ovarian cancers, in which previously available confidence intervals were too wide to be conclusive. Our absolute estimates of population effect emphasised the importance of BMI in driving the incidence of several cancers.

A second important limitation is the potential for residual confounding. Despite adjustment for smoking status (which itself could have been misclassified to some extent[30]), 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[31] (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),[7] breast cancer (by receptor status),[32] non-Hodgkin lymphoma (follicular *vs* diffuse large B-cell),[33] and colon cancer (microsatellite stable *vs* instable,[34–37] or distal *vs* proximal[38,39]). 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.[16] 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

1 WHO. World Health Statistics 2011. Geneva: World Health Organization.
2 Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med* 2003; **348:** 1625–38.
3 Engeland A, Tretli S, Akslen LA, Bjorge T. Body size and thyroid cancer in two million Norwegian men and women. *Br J Cancer* 2006; **95:** 366–70.
4 Engeland A, Tretli S, Bjorge T. Height, body mass index, and ovarian cancer: a follow-up of 1·1 million Norwegian women. *J Natl Cancer Inst* 2003; **95:** 1244–48.
5 Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *BMJ* 2007; **335:** 1134.
6 World Cancer Research Fund, American Institute for Cancer Research. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Washington DC: AICR, 2007.
7 Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 2008; **371:** 569–78.
8 Spencer EA, Appleby PN, Davey GK, Key TJ. Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public Health Nutr* 2002; **5:** 561–65.
9 Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010; **69:** 4–14.
10 Bhaskaran K, Forbes H, Douglas I, Leon DA, Smeeth L. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). *BMJ Open* 2013; **3:** e003389.
11 Cancer Research UK. Cancer incidence for common cancers. 2012. http://info.cancerresearchuk.org/cancerstats/incidence/commoncancers/ (accessed June 20, 2012).
12 Boggon R, van Staa TP, Chapman M, Gallagher AM, Hammad TA, Richards MA. Cancer recording and mortality in the General Practice Research Database and linked cancer registries. *Pharmacoepidemiol Drug Saf* 2012; **22:** 168–75.
13 WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 2004; **363:** 157–63.
14 Her Majesty's Government Department for Communities and Local Government. Index of multiple deprivation. 2011. http://data.gov.uk/dataset/index-of-multiple-deprivation (accessed July 11, 2014).
15 Glymour MM, Greenland S. Causal diagrams. In: Rothman KJ, Greenland S, Lash TL, eds. Modern epidemiology. Philadelphia: Lippincott Williams & Wilkins, 2008.
16 NHS Information Centre. Health Survey for England—2010: trend tables. http://www.ic.nhs.uk/statistics-and-data-collections/health-and-lifestyles-related-surveys/health-survey-for-england/health-survey-for-england--2010-trend-tables (accessed June 19, 2012).
17 Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; **81:** 515–26.

18 Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004; **15:** 615–25.

19 Crosbie EJ, Zwahlen M, Kitchener HC, Egger M, Renehan AG. Body mass index, hormone replacement therapy, and endometrial cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2010; **19:** 3119–30.

20 Song YM, Sung J, Ha M. Obesity and risk of cancer in postmenopausal Korean women. *J Clin Oncol* 2008; **26:** 3395–402.

21 Ulmer H, Bjorge T, Concin H, et al. Metabolic risk factors and cervical cancer in the metabolic syndrome and cancer project (Me-Can). *Gynecol Oncol* 2012; **125:** 330–35.

22 Collaborative Group on Epidemiological Studies of Ovarian Cancer. Ovarian cancer and body size: individual participant meta-analysis including 25,157 women with ovarian cancer from 47 epidemiological studies. *PLoS Med* 2012; **9:** e1001200.

23 Discacciati A, Orsini N, Wolk A. Body mass index and incidence of localized and advanced prostate cancer—a dose-response meta-analysis of prospective studies. *Ann Oncol* 2012; **23:** 1665–71.

24 Smith L, Brinton LA, Spitz MR, et al. Body mass index and risk of lung cancer among never, former, and current smokers. *J Natl Cancer Inst* 2012; **104:** 778–89.

25 Renehan AG, Roberts DL, Dive C. Obesity and cancer: pathophysiological and biological mechanisms. *Arch Physiol Biochem* 2008; **114:** 71–83.

26 The Emerging Risk Factors Collaboration. Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N Engl J Med* 2011; **364:** 829–41.

27 Dregan A, Moller H, Murray-Thomas T, Gulliford MC. Validity of cancer diagnosis in a primary care database compared with linked cancer registrations in England. Population-based cohort study. *Cancer Epidemiol* 2012; **36:** 425–29.

28 Parkinson JP, Davis S, Van Staa T. The General Practice Research Database: now and the future. In: Mann R, Andrews EB, eds. Pharmacovigilance. Chichester: John Wiley and Sons, 2007; 341–48.

29 White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Stat Med* 2010; **29:** 2920–31.

30 Lewis JD, Brensinger C. Agreement between GPRD smoking data: a survey of general practitioners and a population-based survey. *Pharmacoepidemiol Drug Saf* 2004; **13:** 437–41.

31 Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet* 2012; **380:** 219–29.

32 Ritte R, Lukanova A, Berrino F, et al. Adiposity, hormone replacement therapy use and breast cancer risk by age and hormone receptor status: a large prospective cohort study. *Breast Cancer Res* 2012; **14:** R76.

33 Pylypchuk RD, Schouten LJ, Goldbohm RA, Schouten HC, van den Brandt PA. Body mass index, height, and risk of lymphatic malignancies: a prospective cohort study. *Am J Epidemiol* 2009; **170:** 297–307.

34 Campbell PT, Jacobs ET, Ulrich CM, et al. Case-control study of overweight, obesity, and colorectal cancer risk, overall and by tumor microsatellite instability status. *J Natl Cancer Inst* 2010; **102:** 391–400.

35 Hoffmeister M, Blaker H, Kloor M, et al. Body mass index and microsatellite instability in colorectal cancer: a population-based study. *Cancer Epidemiol Biomarkers Prev* 2013; **22:** 2303–11.

36 Hughes LA, Williamson EJ, van Engeland M, et al. Body size and risk for colorectal cancers showing BRAF mutations or microsatellite instability: a pooled analysis. *Int J Epidemiol* 2012; **41:** 1060–72.

37 Slattery ML, Curtin K, Anderson K, et al. Associations between cigarette smoking, lifestyle factors, and microsatellite instability in colon tumors. *J Natl Cancer Inst* 2000; **92:** 1831–36.

38 Hughes LA, Simons CC, van den Brandt PA, Goldbohm RA, van Engeland M, Weijenberg MP. Body size and colorectal cancer risk after 16·3 years of follow-up: an analysis from the Netherlands Cohort Study. *Am J Epidemiol* 2011; **174:** 1127–39.

39 Robsahm TE, Aagnes B, Hjartaker A, Langseth H, Bray FI, Larsen IK. Body mass index, physical activity, and colorectal cancer by anatomical subsites: a systematic review and meta-analysis of cohort studies. *Eur J Cancer Prev* 2013; **22:** 492–505.