THE LANCET

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Bhaskaran K, Douglas I, Forbes H, et al. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. *Lancet* 2014; published online Aug 14. http://dx.doi.org/10.1016/S0140-6736(14)60892-8.

Webappendix Part 1 – additional methods detail and results

1) Identification of cancer cases in the Clinical Practice Research Datalink

To identify cancers in CPRD, the dictionary of Read codes (used by GPs to record clinical diagnoses) was systematically searched to find cancer-related codes using the keywords/word fragments below. The codes picked up by this search were then screened and those indicating malignancy were identified and classified by cancer type (done by KB, reviewed by LS). Each patient's record was then searched for these cancer codes. The earliest code for a particular cancer type was taken as the date of diagnosis. Only the first cancer per patient was considered; patients were then censored at this diagnosis for other cancers. This was due to the difficulty in distinguishing in the data between second de novo cancers and metastates.

Words and word fragments used to search Read code dictionary for cancer-related terms

MELANOMA NEOP TUMOUR CANCER MALIG CARCINOM LEUKAEM METASTA SARCOM LYMPHOM HODGKIN ACROSPIROMA ADAMANTINOMA ADENOACANTHOMA ADENOCARCIONOMA ADENOMA ADENOMATOSIS ANGIOENDOTHELIOMA ANGIOENDOTHELIOMATOSIS ANGIOMYXOMA APUDOMA ARGENTAFFINOMA ARYTHREMIA ASTROCYTOMA BLASTOM BOWEN BURKITT CARCINOID CHEMODECTOMA CHEMOTHERAPY CHLOROMA CHOLANGIOMA CHONDROMATOSIS CHORDOMA CRANIOPHARYNGIOMA CYSTADENOMA DESMOID ECCHONDROSIS EPENDYMOMA EPITHELIOMA ERYTHRAEMIA ERYTHRAEMIA ERYTHRAEMIA ERYTHREMIA ERYTHROPLASIA FIBROMA GAMMOPATHY GASTRINOMA GERMINOMA GLEASON GLIOMA GLUCAGONOMA HAEMANGIOENDOTHELIOMA HAEMANGIOENDOTHELIOMA HEMANGIOENDOTHELIOMA HEPATOMA HISTIOCYTIC HISTIOCYTOMA HISTIOCYTOSIS HYDATIDIFORM HYPERNEPHR HYPERNEPHR IMMUNOPROLIFERATIVE IMMUNOPROLIFERATIVE INSULINOMA KAHLER LEIOMYOMATOSIS LETTERER LYMPHANGIOMYOMATOSIS LYMPHOM LYMPHOPROLIFERATIVE MASTOCYTOMA MASTOCYTOSIS MECKEL MENINGIOMA MESENCHYMOMA MESONEPHROMA MESOTHELIOMA MESOTHELIOMA MYELODYSPLASTIC MYELOFIBROSIS MYELOMA MYELOMA MYELOPROLIFERATIVE MYELOSCLEROSIS MYELOSIS NEPHROMA NEURILEMMOMA NEURINOMATOSIS NEUROCYTOMA NEUROFIBROMATOSIS OSTEOCLASTOMA PAGET PANCOAST PANMYELOSIS PARAGANGLIOMA PERICYTOMA PINEALOMA PINEOCYTOMA PLASMACYTOMA PLASTICA POLYCYTHAEMIA POLYCYTHEMIA POLYEMBRYOMA PSEUDOMYXOMA RADIOTHERAPY SEMINOMA SEZARY TERATOMA TERATOMA THECOMA THROMBOCYTHAEMIA THROMBOCYTHEMIA THYMOMA VIPOMA WALDENSTROM [M] "ANGIOIMMUNOBLASTIC LYMPHADENOPATHY" "ATYPICAL FIBROXANTHOMA" BRILL CA CA-IN-SITU "DI GUGLIELMO" "GIANT PIGMENTED NAEVUS" "GIANT PIGMENTED NEVUS" "HEAVY CHAIN" "HUTCHINSON'S MELANOTIC" "MAST CELL" "MYCOSIS FUNGOIDES" "NEO/" "REFRACTORY ANAEMIA" "REFRACTORY ANEMIA" "RODENT ULCER" "STROMAL MYOSIS" "STRUMA OVARII" "TRANSITIONAL CELL PAPILLOMA, INVERTED" "UROTHELIAL PAPILLOMA"

2) Specification of restricted cubic spline models relating BMI to cancer

In order to investigate possible non-linearity in the BMI-cancer association, BMI was included in our adjusted models as a restricted cubic spline. Knots were placed at equal percentiles of the data. The optimal number of knots was determined as follows: for each cancer, we fit models with 3, 4, 5, 6, and 7 knots, and for each of these we calculated the Akaike Information Criterion (AIC, which balances model fit against parsimony); the number of knots resulting in the minimum AIC was then chosen. Since this procedure was carried out for each cancer outcome separately, the number of knots varies across the cancer-specific models. However, for each cancer type, we retained the same number of knots when subsequently introducing interaction terms to assess effect modification.

3) Calculating population attributable fraction for overweight/obese, accounting for different risk in the underweight

Let :

p_{uw} = prevalence of underweight
p_{nw} = prevalence of normal weight
p_{ow} = prevalence of overweight including obesity
RR_{uw} = relative risk of cancer for underweight vs normal weight
RR_{ow} = relative risk of cancer for overweight vs normal weight

and let C be a constant such that the number of cancers under current overweight/obesity levels

$$= C*(p_{uw}*RR_{uw} + p_{nw}*1 + p_{ow}*RR_{ow})$$

Then (assuming causality), the number of cancers in absence of overweight/obesity

$$= C*(p_{uw}*RR_{uw} + (p_{nw} + p_{ow})*1)$$

and the population attributable fraction (PAF)

$$= (C*(p_{uw}*RR_{uw} + p_{nw}*1 + p_{ow}*RR_{ow}) - C*(p_{uw}*RR_{uw} + (p_{nw} + p_{ow})*1)) / C*(p_{uw}*RR_{uw} + p_{nw}*1 + p_{ow}*RR_{ow}$$

$$= p_{ow}*(RR_{ow} - 1) / (p_{uw}*RR_{uw} + p_{nw}*1 + p_{ow}*RR_{ow})$$

For our calculations, we fitted 3-category BMI models to estimate RR_{uw} and RR_{ow} (separately for men and women) for each cancer site. p_{uw}, p_{nw} and p_{ow} were obtained from published Health Survey for England 2010 Trend Tables, as cited in the manuscript.

Table W1.1: Characteristics of individuals excluded from the study

Characteristics (cell contents are N and % unless otherwise stated)	
N	4,794,834
Person-years of follow-up in CPRD	
Mean (sd)	5.8 (5.7)
Median	3.8
IQR	(0.9-9.7)
Total p-years (millions)	27.795
Age at start of CPRD follow-up (yrs)	
Median (IQR)	32.8 (21.5-54.8)
Gender	
Male	2,365,399 (49.3)
Female	2,428,509 (50.7)
Missing	926 (0.02)
Smoking status (first recorded)	
Never-smoker	1,259,326 (26.3)
Current smoker	804,939 (16.8)
Ex-smoker	282,465 (5.9)
Missing	2,448,104 (51.1)
Alcohol use (first recorded)	
Non-drinker	352,103 (7.3)
Current drinker	498,169 (10.4)
Ex-drinker	50,757 (1.1)
Missing	3,893,805 (81.2)
Index of multiple deprivation quintile	
1 (least deprived)	856,697 (17.9)
2	913,090 (19.0)
3	1,044,813 (21.8)
4	1,142,878 (23.8)
5 (most deprived)	837,356 (17.5)
Calendar year at start of CPRD follow-up	
≤1989	237,561 (5.0)
1990-94	669,807 (14.0)
1995-99	1,034,277 (21.6)
2000-04	1,201,002 (25.0)
2005-09	979,500 (20.4)
≥2010	672,618 (14.0)
Any record of cancer	231,764 (4.8)
Prior to CPRD follow-up	139,598 (2.9)
During CPRD follow-up	91,442 (1.9)

Table W1.2: Events, person-years and relative risks for each cancer by BMI category, adjusted for age and sex only

cell contents: events/100,000 p-yrs RR (99% CI)	BMI category (BMI ranges in kg/m^2)					
	Underweight (<18.5)	**Normal weight (18.5-24.9)**	**Overweight (25.0-29.9)**	**Moderately obese (30.0-34.9)**	**Severely obese (≥35.0)**	**Total**
	121/8.4	*2907/159.0*	*3491/113.8*	*1147/40.3*	*310/16.8*	*7976/338.4*
Oral cavity (C00-06)	2.52 (1.69, 3.76)	1.00 (REF)	0.78 (0.65, 0.93)	0.81 (0.63, 1.03)	0.82 (0.57, 1.20)	-
	145/8.4	*1870/159.0*	*2126/113.8*	*810/40.3*	*262/16.8*	*5213/338.4*
Oesophagus (C15)	2.17 (1.74, 2.72)	1.00 (REF)	1.01 (0.93, 1.10)	1.18 (1.06, 1.31)	1.31 (1.10, 1.55)	-
	68/8.4	*1230/159.0*	*1384/113.8*	*515/40.3*	*140/16.8*	*3337/338.4*
Stomach (C16)	1.42 (1.03, 1.96)	1.00 (REF)	1.03 (0.93, 1.14)	1.18 (1.03, 1.36)	1.12 (0.89, 1.41)	-
	204/8.4	*4685/159.0*	*5671/113.8*	*2181/40.3*	*724/16.8*	*13465/338.4*
Colon (C18)	1.08 (0.90, 1.30)	1.00 (REF)	1.16 (1.10, 1.22)	1.32 (1.23, 1.41)	1.36 (1.23, 1.51)	-
	92/8.4	*2262/159.0*	*2567/113.8*	*898/40.3*	*304/16.8*	*6123/338.4*
Rectum (C20)	1.10 (0.83, 1.45)	1.00 (REF)	1.04 (0.97, 1.13)	1.09 (0.98, 1.21)	1.18 (1.01, 1.38)	-
	28/8.4	*567/159.0*	*773/113.8*	*346/40.3*	*145/16.8*	*1859/338.4*
Liver (C22)	1.35 (0.82, 2.22)	1.00 (REF)	1.25 (1.08, 1.44)	1.71 (1.43, 2.04)	2.38 (1.87, 3.03)	-
	2/8.4	*95/159.0*	*125/113.8*	*57/40.3*	*24/16.8*	*303/338.4*
Gall bladder (C23)	0.47 (0.07, 2.99)	1.00 (REF)	1.37 (0.97, 1.96)	1.77 (1.15, 2.73)	2.11 (1.16, 3.81)	-
	65/8.4	*1478/159.0*	*1502/113.8*	*583/40.3*	*223/16.8*	*3851/338.4*
Pancreas (C25)	1.06 (0.76, 1.47)	1.00 (REF)	0.98 (0.89, 1.07)	1.10 (0.97, 1.25)	1.29 (1.07, 1.56)	-
	682/8.4	*8626/159.0*	*7057/113.8*	*2312/40.3*	*662/16.8*	*19339/338.4*
Lung (C34)	2.15 (1.94, 2.38)	1.00 (REF)	0.74 (0.71, 0.78)	0.72 (0.68, 0.77)	0.66 (0.59, 0.73)	-
	125/8.4	*3720/159.0*	*3170/113.8*	*1114/40.3*	*376/16.8*	*8505/338.4*
Mal melanoma (C43)	0.79 (0.62, 1.00)	1.00 (REF)	1.00 (0.94, 1.06)	0.99 (0.91, 1.08)	0.86 (0.75, 0.99)	-
	149/4.1	*3770/54.5*	*1537/21.0*	*554/8.3*	*288/5.0*	*6298/93.0*
Breast- premeno (C50)	0.82 (0.66, 1.02)	1.00 (REF)	0.90 (0.83, 0.97)	0.78 (0.70, 0.88)	0.64 (0.55, 0.75)	-
	416/1.9	*11838/41.5*	*9948/31.7*	*4227/13.2*	*1980/6.3*	*28409/94.7*
Breast – postmeno (C50)	0.77 (0.68, 0.88)	1.00 (REF)	1.08 (1.04, 1.12)	1.11 (1.06, 1.16)	1.10 (1.03, 1.17)	-
	45/5.6	*652/86.9*	*392/44.9*	*188/18.0*	*112/9.7*	*1389/165.1*
Cervix (C53)	1.15 (0.77, 1.71)	1.00 (REF)	1.10 (0.93, 1.30)	1.32 (1.06, 1.63)	1.49 (1.15, 1.95)	-
	31/5.6	*723/86.9*	*819/44.9*	*577/18.0*	*608/9.7*	*2758/165.1*
Uterus (C54-55)	0.93 (0.58, 1.50)	1.00 (REF)	1.52 (1.33, 1.74)	2.65 (2.29, 3.06)	5.86 (5.08, 6.76)	-
	72/5.9	*1541/94.0*	*1231/51.0*	*548/20.7*	*292/11.0*	*3684/182.6*
Ovaries (C56)	0.97 (0.71, 1.32)	1.00 (REF)	1.08 (0.98, 1.20)	1.17 (1.03, 1.33)	1.30 (1.10, 1.54)	-
	168/2.4	*8774/62.9*	*12119/61.1*	*3263/18.8*	*577/5.5*	*24901/150.8*
Prostate (C61)	0.79 (0.65, 0.97)	1.00 (REF)	1.06 (1.02, 1.10)	0.97 (0.92, 1.02)	0.78 (0.70, 0.87)	-
	18/8.4	*623/159.0*	*786/113.8*	*334/40.3*	*145/16.8*	*1906/338.4*
Kidney (C64)	0.82 (0.44, 1.52)	1.00 (REF)	1.18 (1.03, 1.36)	1.48 (1.24, 1.77)	1.99 (1.56, 2.52)	-
	121/8.4	*2907/159.0*	*3491/113.8*	*1147/40.3*	*310/16.8*	*7976/338.4*
Bladder (C67)	1.22 (0.96, 1.55)	1.00 (REF)	1.05 (0.98, 1.12)	1.09 (0.99, 1.19)	1.07 (0.92, 1.25)	-
	39/8.4	*1195/159.0*	*1159/113.8*	*419/40.3*	*162/16.8*	*2974/338.4*
Brain/CNS (C71-72)	0.85 (0.56, 1.30)	1.00 (REF)	1.01 (0.90, 1.12)	1.05 (0.91, 1.22)	1.15 (0.92, 1.42)	-
	15/8.4	*432/159.0*	*303/113.8*	*122/40.3*	*69/16.8*	*941/338.4*
Thyroid (C73)	0.68 (0.35, 1.34)	1.00 (REF)	1.05 (0.86, 1.28)	1.11 (0.85, 1.46)	1.37 (0.98, 1.91)	-
	97/8.4	*2789/159.0*	*2751/113.8*	*950/40.3*	*359/16.8*	*6946/338.4*
NHL (C82-85)	0.90 (0.69, 1.17)	1.00 (REF)	1.00 (0.93, 1.07)	0.99 (0.90, 1.10)	1.08 (0.93, 1.25)	-
	44/8.4	*1114/159.0*	*1233/113.8*	*419/40.3*	*159/16.8*	*2969/338.4*
Mult myeloma (C90)	1.01 (0.68, 1.50)	1.00 (REF)	1.05 (0.94, 1.17)	1.03 (0.89, 1.20)	1.20 (0.96, 1.49)	-
	73/8.4	*2185/159.0*	*2394/113.8*	*871/40.3*	*310/16.8*	*5833/338.4*
Leukemia (C91-95)	0.83 (0.61, 1.14)	1.00 (REF)	1.07 (0.99, 1.15)	1.16 (1.04, 1.29)	1.28 (1.10, 1.50)	-

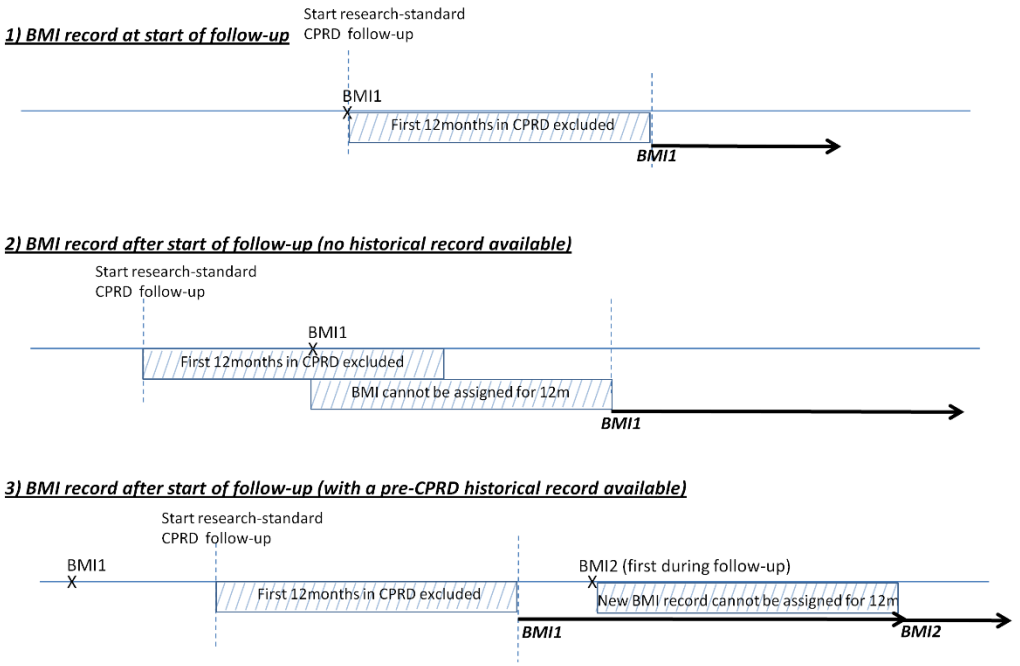
CNS=Central nervous system (inc brain); NHL = non-Hodgkin lymphoma

Table W1.3: Estimated linear BMI-cancer association (HR per 5kg/m2 increase in BMI) across sensitivity analyses (numerical results)

	Estimated linear effect (% increase per 5kg/m^2) and 99% CI							
	Original analysis	1) Re-instate follow-up just after BMI	2) Exclude 3y follow-up after BMI	3) Exclude if BMI>12m from registr'n	4) Drop pre-research standard follow-up BMIs	5) Using linked outcome data	6) Adjust for GP "attender" status	7) Adjust non-proportional hazards
N patients	*5,243,978*	*5,250,073*	*1,200,028*	*2,575,869*	*4,430,263*	*2,968,114*	*5,243,978*	*5,243,978*
Oral cavity (C00-06)	0.81 (0.74, 0.89)	0.80 (0.73, 0.88)	0.90 (0.75, 1.09)	0.77 (0.63, 0.92)	0.77 (0.69, 0.85)	0.85 (0.78, 0.92)	0.81 (0.74, 0.89)	N/A
Oesophagus (C15)	1.03 (0.99, 1.08)	1.02 (0.97, 1.06)	1.03 (0.94, 1.13)	1.04 (0.95, 1.14)	1.03 (0.98, 1.08)	1.04 (0.99, 1.09)	1.03 (0.99, 1.07)	1.03 (0.99, 1.07)
Stomach (C16)	1.03 (0.98, 1.09)	1.01 (0.96, 1.07)	1.06 (0.93, 1.20)	1.04 (0.93, 1.17)	1.02 (0.96, 1.08)	1.06 (1.00, 1.12)	1.03 (0.97, 1.09)	N/A
Colon (C18)	1.10 (1.07, 1.13)	1.08 (1.06, 1.11)	1.12 (1.06, 1.18)	1.10 (1.04, 1.16)	1.10 (1.07, 1.13)	1.09 (1.06, 1.12)	1.10 (1.07, 1.12)	N/A
Rectum (C20)	1.04 (1.00, 1.08)	1.03 (0.99, 1.07)	1.05 (0.97, 1.15)	1.01 (0.93, 1.09)	1.05 (1.00, 1.09)	1.02 (0.97, 1.06)	1.04 (1.00, 1.08)	N/A
Liver (C22)	1.19 (1.12, 1.27)	1.18 (1.11, 1.26)	1.32 (1.15, 1.51)	1.09 (0.95, 1.25)	1.19 (1.11, 1.27)	1.23 (1.15, 1.31)	1.19 (1.12, 1.27)	1.19 (1.12, 1.27)
Gall bladder (C23)	1.31 (1.12, 1.52)	1.31 (1.13, 1.53)	1.31 (0.93, 1.84)	1.22 (0.88, 1.69)	1.35 (1.15, 1.59)	1.38 (1.19, 1.61)	1.31 (1.13, 1.53)	N/A
Pancreas (C25)	1.05 (1.00, 1.10)	1.04 (0.99, 1.09)	1.10 (0.99, 1.21)	1.04 (0.95, 1.15)	1.05 (0.99, 1.10)	1.05 (1.00, 1.11)	1.05 (1.00, 1.10)	1.05 (1.00, 1.10)
Lung (C34)	0.82 (0.81, 0.84)	0.82 (0.80, 0.84)	0.84 (0.80, 0.88)	0.82 (0.78, 0.86)	0.82 (0.80, 0.84)	0.83 (0.81, 0.85)	0.82 (0.80, 0.84)	0.82 (0.80, 0.84)
Melanoma (C43)	0.99 (0.96, 1.02)	0.99 (0.96, 1.02)	0.97 (0.90, 1.03)	0.99 (0.93, 1.05)	1.00 (0.96, 1.03)	1.00 (0.97, 1.05)	0.99 (0.96, 1.02)	0.99 (0.96, 1.02)
Breast –pre (C50)	0.89 (0.86, 0.92)	0.89 (0.86, 0.92)	0.88 (0.82, 0.96)	0.91 (0.86, 0.97)	0.89 (0.86, 0.92)	0.92 (0.88, 0.96)	0.89 (0.86, 0.92)	N/A
Breast –post (C50)	1.05 (1.03, 1.07)	1.05 (1.03, 1.06)	1.05 (1.02, 1.09)	1.05 (1.02, 1.09)	1.06 (1.04, 1.07)	1.05 (1.03, 1.07)	1.05 (1.03, 1.07)	1.05 (1.03, 1.07)
Cervix (C53)	1.10 (1.03, 1.17)	1.09 (1.02, 1.16)	1.11 (0.95, 1.28)	1.15 (1.03, 1.30)	1.09 (1.01, 1.17)	1.07 (0.99, 1.16)	1.10 (1.03, 1.17)	N/A
Uterus (C54-55)	1.62 (1.56, 1.69)	1.63 (1.56, 1.69)	1.69 (1.56, 1.84)	1.65 (1.52, 1.79)	1.62 (1.56, 1.69)	1.62 (1.56, 1.68)	1.62 (1.56, 1.69)	N/A
Ovaries (C56)	1.09 (1.04, 1.14)	1.09 (1.04, 1.13)	1.11 (1.01, 1.21)	1.05 (0.96, 1.14)	1.09 (1.04, 1.14)	1.12 (1.07, 1.17)	1.09 (1.04, 1.14)	N/A
Prostate (C61)	0.98 (0.95, 1.00)	0.97 (0.95, 0.99)	0.96 (0.91, 1.00)	1.01 (0.96, 1.05)	0.98 (0.95, 1.00)	0.98 (0.96, 1.01)	0.97 (0.95, 1.00)	0.97 (0.95, 1.00)
Kidney (C64)	1.25 (1.17, 1.33)	1.24 (1.16, 1.32)	1.20 (1.04, 1.39)	1.22 (1.08, 1.39)	1.25 (1.16, 1.33)	1.21 (1.15, 1.28)	1.24 (1.17, 1.32)	N/A
Bladder (C67)	1.03 (0.99, 1.06)	1.02 (0.99, 1.06)	0.99 (0.91, 1.07)	1.00 (0.93, 1.08)	1.03 (0.99, 1.07)	1.03 (1.00, 1.07)	1.02 (0.99, 1.06)	1.03 (0.99, 1.06)
Brain/CNS (C71-72)	1.04 (0.99, 1.10)	1.04 (0.99, 1.10)	1.04 (0.92, 1.16)	1.08 (0.97, 1.19)	1.04 (0.98, 1.10)	1.03 (0.97, 1.10)	1.04 (0.98, 1.10)	N/A
Thyroid (C73)	1.09 (1.00, 1.19)	1.08 (0.99, 1.17)	1.09 (0.90, 1.32)	1.02 (0.87, 1.18)	1.12 (1.02, 1.23)	1.09 (0.99, 1.19)	1.09 (1.00, 1.18)	N/A
NHL (C82-85)	1.03 (0.99, 1.06)	1.02 (0.98, 1.05)	1.08 (1.00, 1.16)	0.99 (0.92, 1.06)	1.03 (0.99, 1.07)	1.03 (0.99, 1.07)	1.03 (0.99, 1.06)	1.03 (0.99, 1.06)
Myeloma (C90)	1.03 (0.98, 1.09)	1.02 (0.97, 1.08)	1.07 (0.96, 1.20)	1.07 (0.95, 1.20)	1.01 (0.95, 1.07)	1.03 (0.96, 1.10)	1.03 (0.97, 1.09)	N/A
Leukemia (C91-95)	1.09 (1.05, 1.13)	1.08 (1.04, 1.12)	1.09 (1.01, 1.19)	1.08 (1.00, 1.17)	1.09 (1.04, 1.13)	1.06 (1.01, 1.11)	1.09 (1.05, 1.13)	N/A

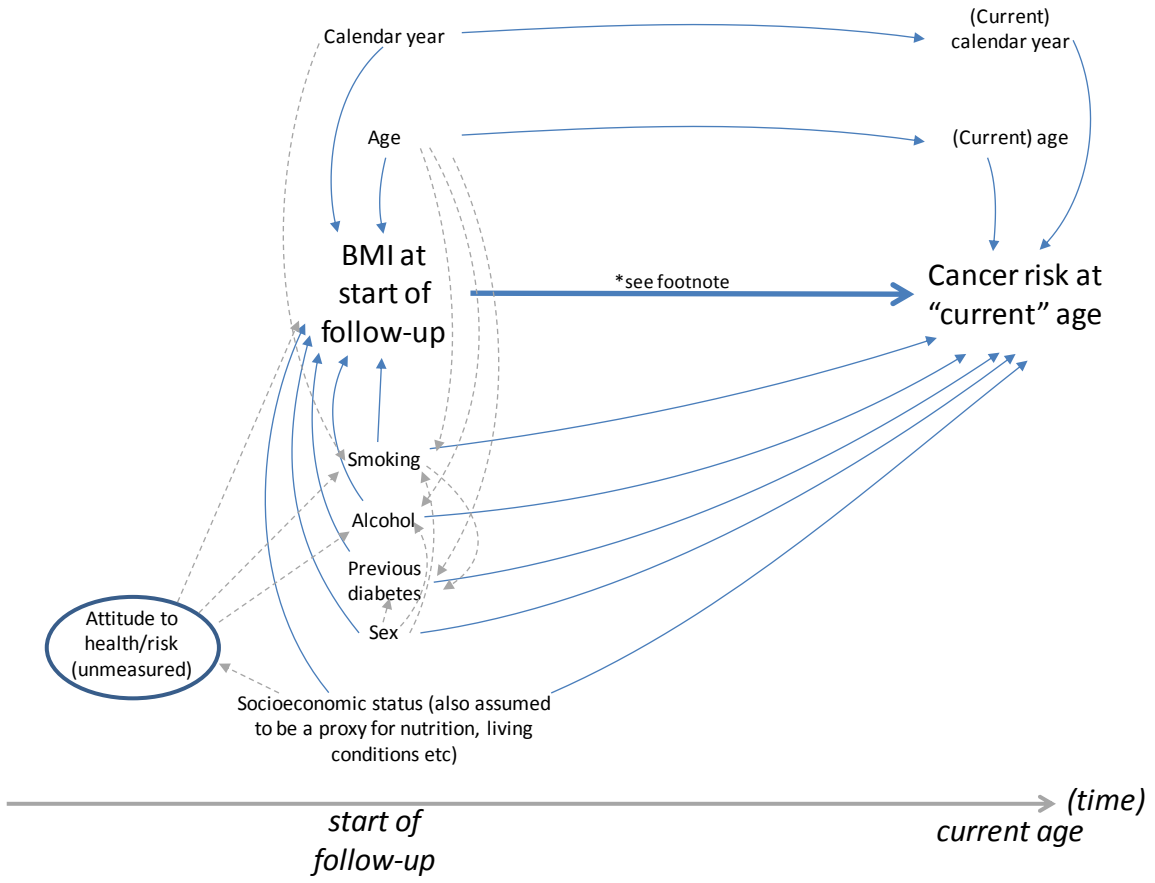
Notes: See also Figure W1.8 for graphical presentation of results. % increase in risk calculated as (1-hazard ratio)x100%; separate model for each cancer with linear BMI term, all models adjusted for adjusted for age, diabetes status, smoking, alcohol, socioeconomic status, calendar year, and stratified by sex. An additional sensitivity analysis for liver cancer in which missing alcohol status was dealt with using multiple imputation had little effect on the estimated effect of BMI (19.8% increase in risk per 5kg/m^2 increase in BMI, 99% CI 13.5 to 26.2).

Figure W1.1: Assignment of exposure in 3 example scenarios



Notes: Exposure assignment is shown by the bold arrows.
Scenario (1) – individual enters the risk set 12 months after start of follow-up because there is a BMI measure available at start of follow-up (BMI1). The patient's exposure is assigned as BMI1 from this point on.
Scenario (2) - individual has a BMI some months after start of follow-up in CPRD. No pre-CPRD historical BMI data are available. Individual therefore enters the risk set 12m after the first BMI recording (to guard against reverse causality).
Scenario (3) – individual has a BMI some time after start of follow-up in CPRD, but a historical (pre-follow-up) BMI record is available. The individual therefore enters the risk set 12 months after start of follow-up in CPRD with the most recent pre-CPRD BMI record assigned as exposure. Once the first BMI recorded during CPRD follow-up becomes available (i.e. 12 months after its recording), the exposure is updated to use this measure.

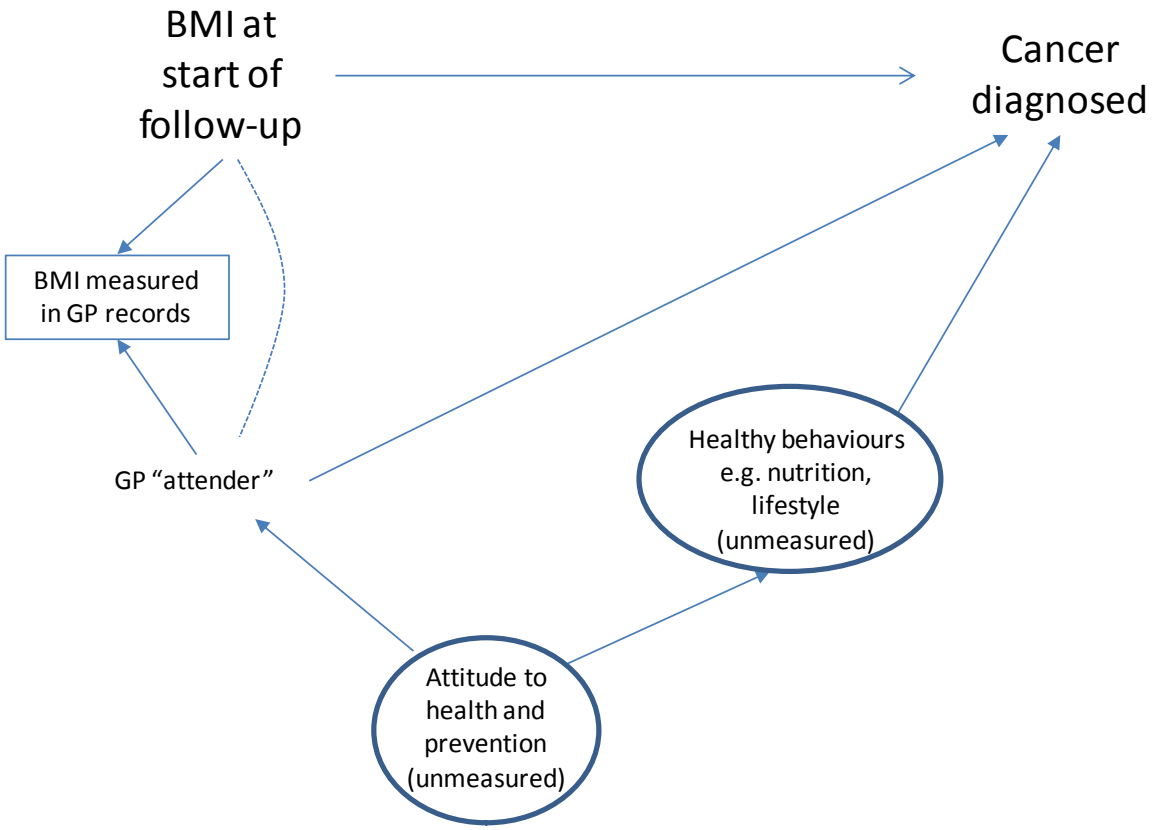
Figure W1.2: Simplified directed acyclic graph (DAG) illustrating implicitly assumed causal structure underlying our adjusted models.



* Causal path of interest representing total causal effect, including through mediators (such as diabetes during follow-up, BMI during follow-up)

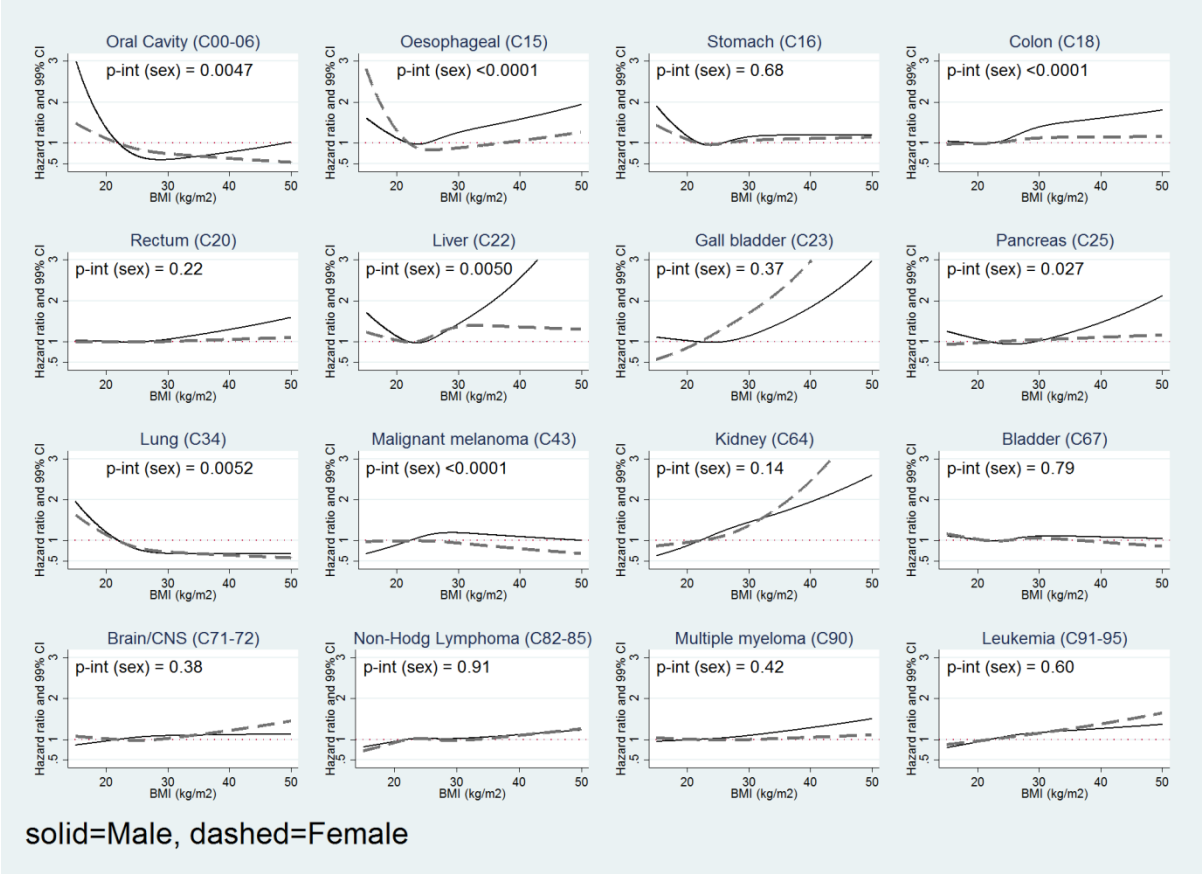
Note: under this assumed causal framework, adjustment for all measured variables in the left of the diagram (as per our adjusted models stratified by sex) blocks all confounding paths. Mediators along causal pathways (e.g. between sex and cancer risk) are generally omitted from the diagram for simplicity. Some paths that are not immediately between exposure and outcome (i.e. paths between "confounders") have been formatted as dashed lines; this is purely to aid clarity.

Figure W1.3: Simplified directed acyclic graph (DAG) illustrating possible selection (collider stratification) bias due to restriction to individuals with BMI measured (motivation for sensitivity analysis #6)



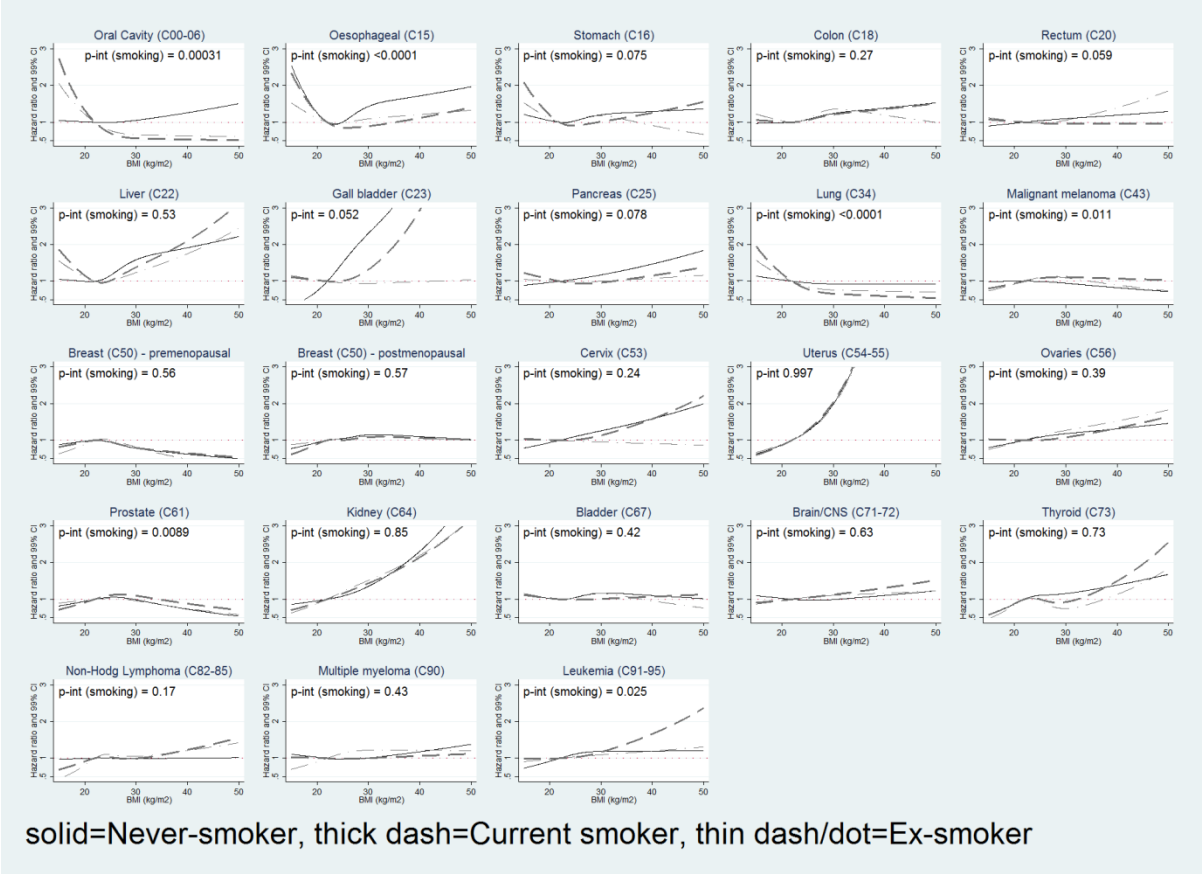
Notes: Under the assumed mechanism illustrated above, conditioning on BMI having been measured (our main inclusion criterion) induces association between BMI and GP attendance (dotted line), which opens up biasing paths between exposure (BMI) and outcome (cancer). Adjustment for GP attendance (as in sensitivity analysis #6) would block such biasing paths.

Figure W1.4: Association between BMI and specific cancers: effect modification by gender



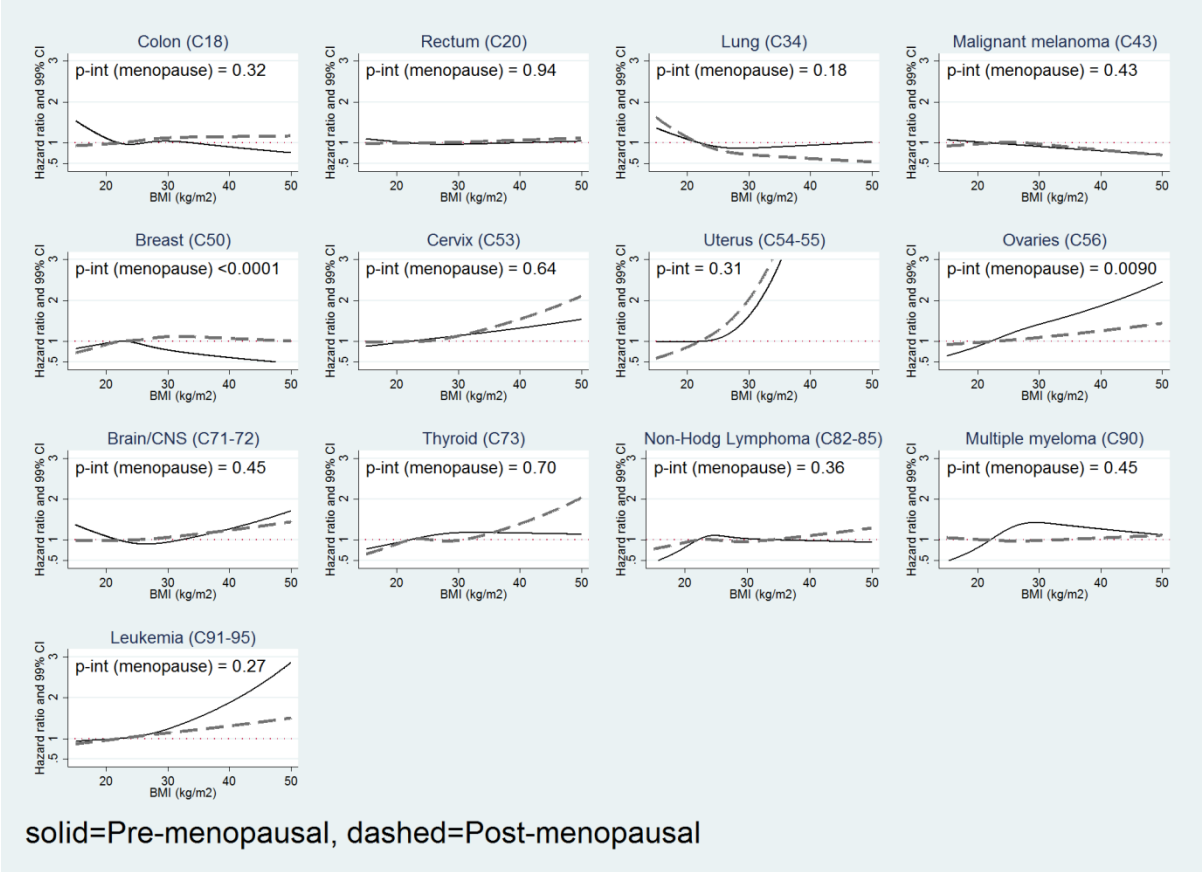
Note: separate model for each cancer with a 3 knot restricted cubic spline for BMI (knots placed at equal percentiles of BMI), adjusted for age, calendar year, diabetes status, alcohol use, smoking (all at time of BMI recording), socioeconomic status (Index of Multiple Deprivation), and with interaction terms between the BMI spline basis variables and gender. Curves show hazard ratios compared with the chosen reference BMI of 22kg/m^2.
Thyroid cancer is omitted from the figure as there were insufficient events to obtain reliable estimated curves.
In a post-hoc analysis restricted to never-smokers, there was no evidence of effect modification by gender for oral cavity and lung cancers (p-interaction 0.42 and 0.55 respectively)
p-int = p-value for interaction

Figure W1.5: Association between BMI and specific cancers: effect modification by smoking status



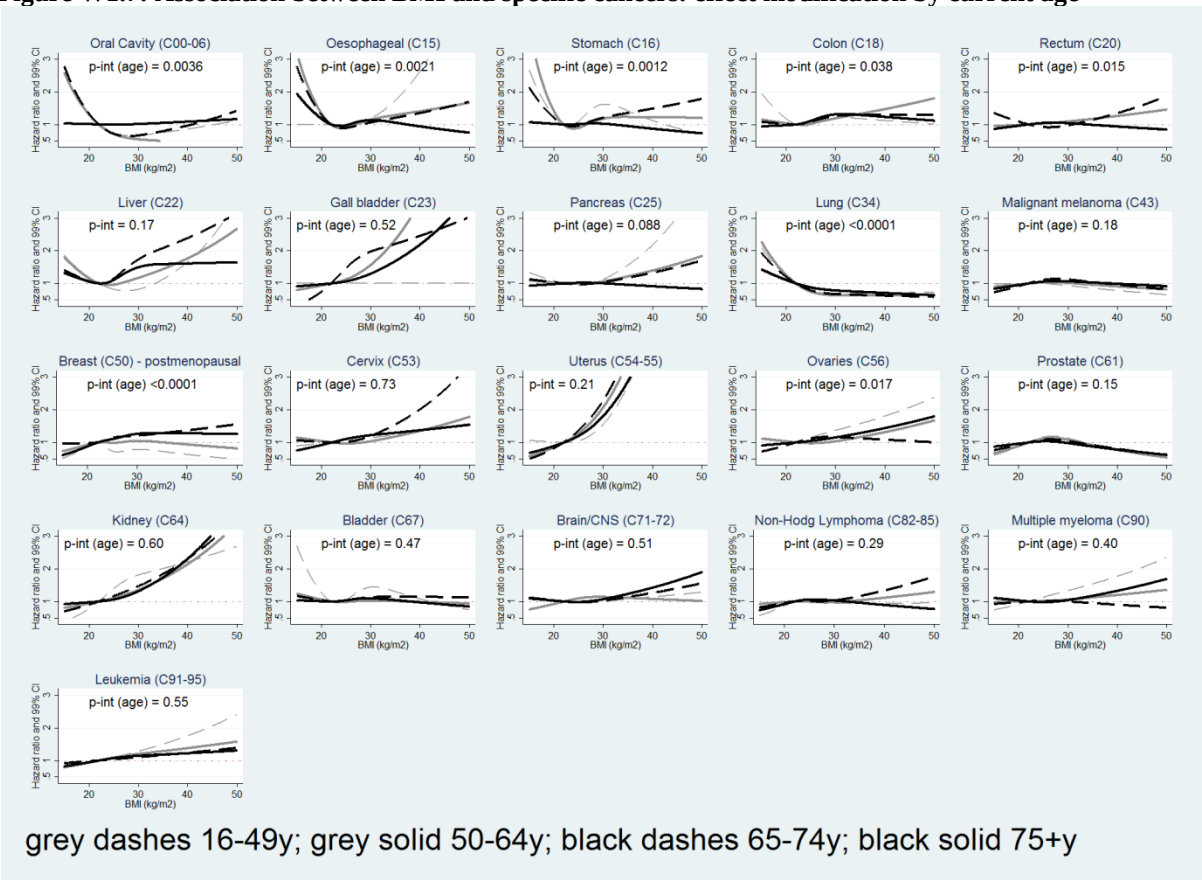
Note: separate model for each cancer with a 3 knot restricted cubic spline for BMI (knots placed at equal percentiles of BMI), stratified by gender and adjusted for age, calendar year, diabetes status, alcohol use, smoking (all at time of BMI recording), socioeconomic status (Index of Multiple Deprivation), and with interaction terms between the BMI spline basis variables and smoking. Curves show hazard ratios compared with the chosen reference BMI of 22kg/m^2.
p-int = p-value for interaction

Figure W1.6: Association between BMI and specific cancers in women: effect modification by menopausal status



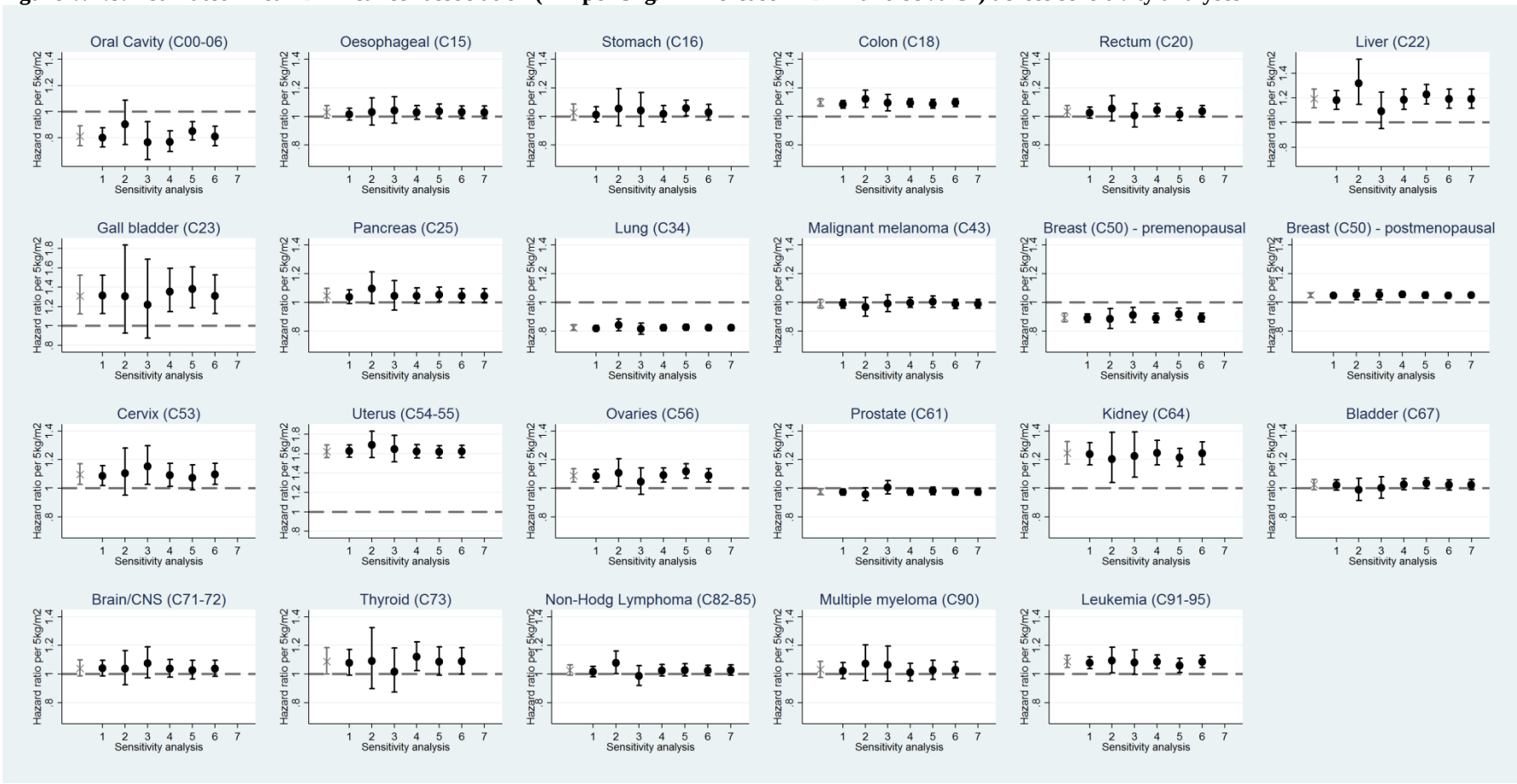
Note: separate model for each cancer with a 3 knot restricted cubic spline for BMI (knots placed at equal percentiles of BMI), adjusted for age, calendar year, diabetes status, alcohol use, smoking (all at time of BMI recording), socioeconomic status (Index of Multiple Deprivation), and with interaction terms between the BMI spline basis variables and menopausal status; men omitted from model. Curves show hazard ratios compared with the chosen reference BMI of 22kg/m^2.
Oral, oesophageal, stomach, liver, gall bladder, pancreas, kidney, and bladder are omitted from the figure as there were insufficient pre-menopause events to obtain reliable estimated curves.
p-int = p-value for interaction

Figure W1.7: Association between BMI and specific cancers: effect modification by current age



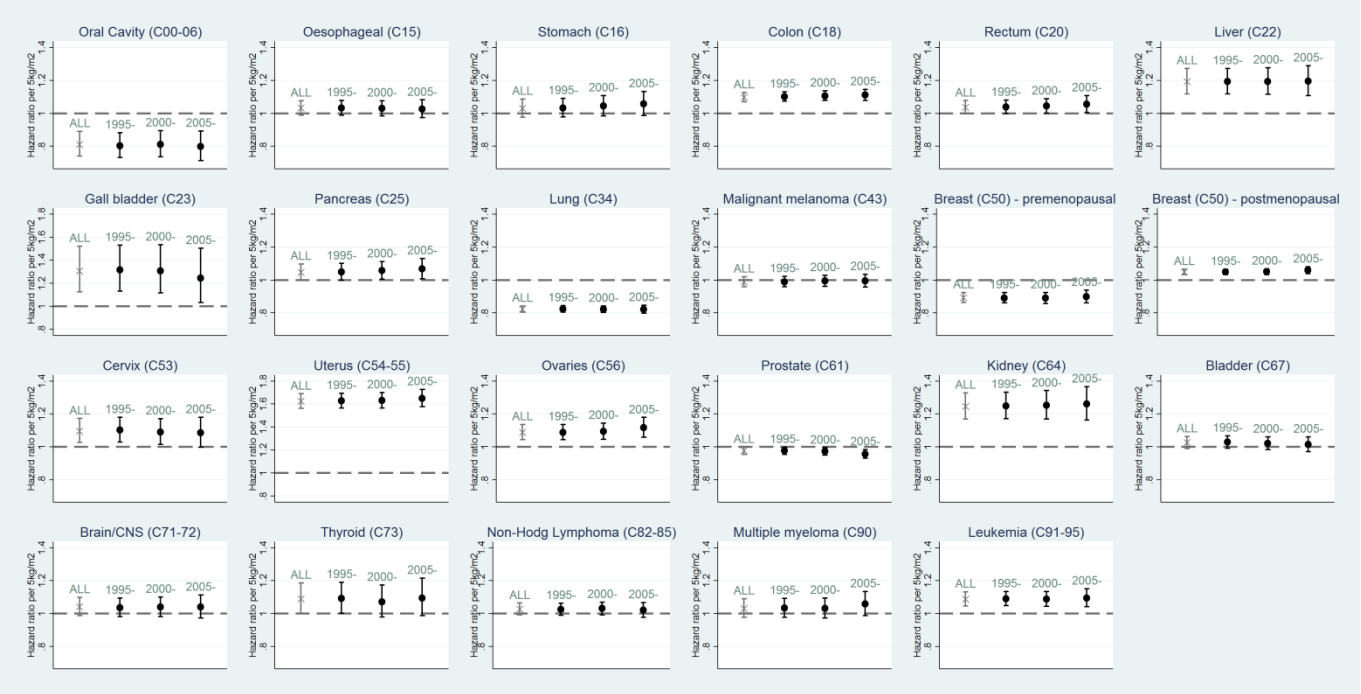
Note: separate model for each cancer with a 3 knot restricted cubic spline for BMI (knots placed at equal percentiles of BMI), stratified by gender and adjusted for age, calendar year, diabetes status, alcohol use, smoking (all at time of BMI recording), socioeconomic status (Index of Multiple Deprivation), and with interaction terms between the BMI spline basis variables and time-updated age group. Curves show hazard ratios compared with the chosen reference BMI of 22kg/m^2.
The age groups were chosen to divide the number of cancer events into approximate quartiles. For gall bladder cancer, there were insufficient events in the 16-49 years age group so this was combined with the 50-64 age group. Pre-menopausal breast cancer was not included since by our definition of pre-menopausal, all such individuals were in the single 16-49 age group.
Thyroid cancer is omitted from the figure as there were insufficient events to obtain reliable estimated curves.In a post-hoc analysis restricted to never-smokers, there was no evidence of effect modification by age for oral cavity, stomach and lung cancers (p-interaction 0.65, 0.11 and 0.40 respectively)
p-int = p-value for interaction

Figure W1.8: Estimated linear BMI-cancer association (HR per 5kg/m2 increase in BMI and 99% CI) across sensitivity analyses



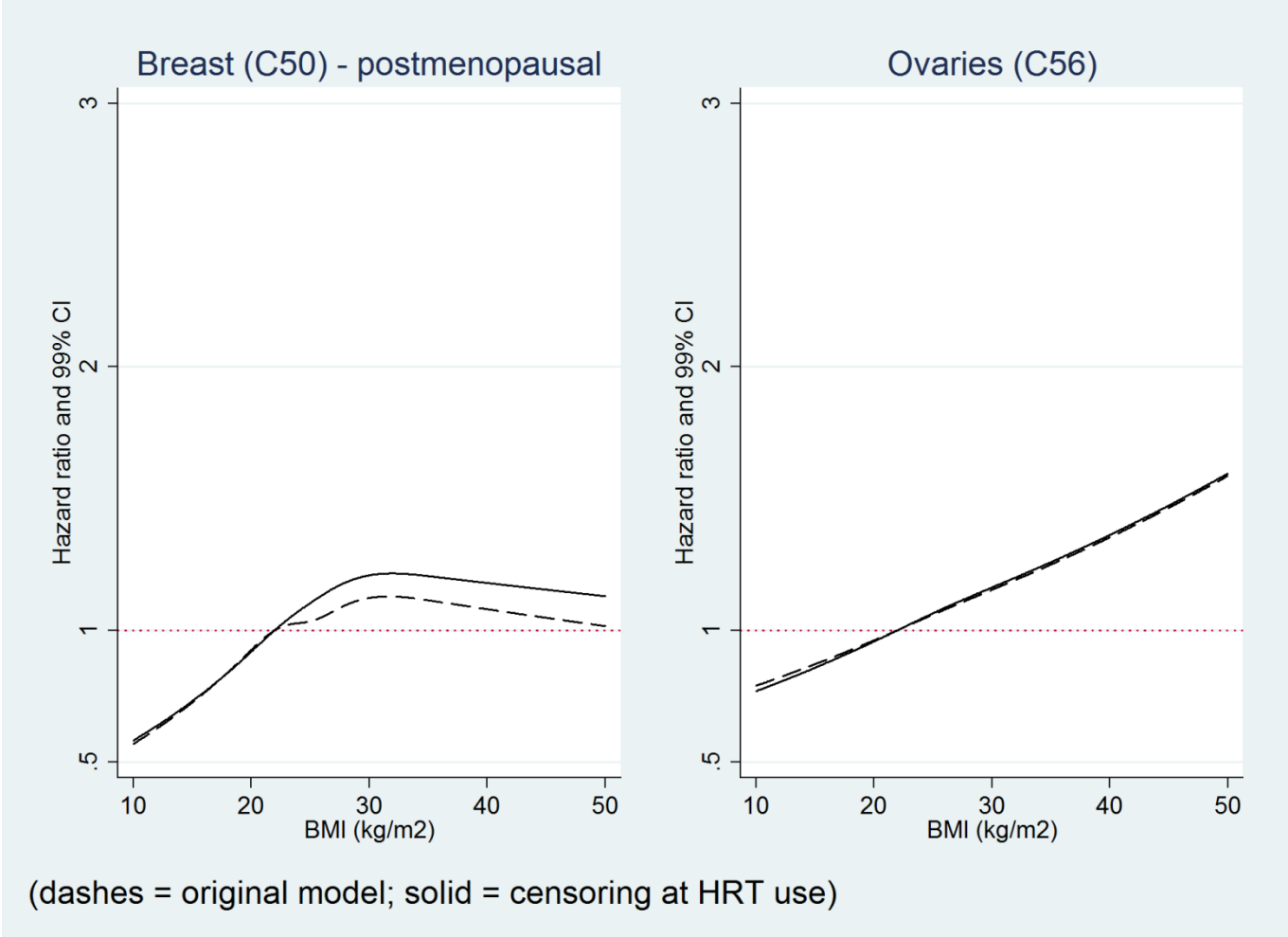
Notes: See also Table W2.1 for numerical results. Left-hand estimate in each panel (in grey) shows effect from the main analysis. Sensitivity analyses (1) to (4) were as follows:
(1) Including instead of excluding the follow-up in the 12 months after a BMI record;
(2) Extending the exclusion period after a BMI record to 3 years;
(3) Restricting the analysis to those with a BMI measured within 12 months of their GP registration;
(4) Using only BMIs recorded during CPRD follow-up (i.e. "up to research standard")
(5) Using linked hospital data and death certificates in addition to CPRD to identify cancers (analysis limited to those in practices participating in a data linkage scheme);
(6) Adjusting for whether patient consulted in the first 12 months of CPRD follow-up, to adjust for potential selection bias mechanisms
(7) (Where applicable) adding interaction terms between current age (time-updated) and any variables in the model for which there was evidence of violation of proportional hazards, in order to properly account for non-proportional hazards (the following interactions terms were added for specific cancers: bladder – smoking, alcohol; postmenopausal breast – smoking and diabetes; liver – alcohol; lung - smoking, age at BMI, index of multiple deprivation; malignant melanoma – alcohol; non-Hodgkin lymphoma – alcohol; oesophageal – smoking, diabetes status; pancreas – smoking, alcohol; prostate; smoking, diabetes, calendar year; no evidence of non-proportional hazards for cervical, brain/CNS, colon, rectal, gall bladder, gastric, kidney, multiple myeloma, oral cavity, ovarian, thyroid, uterus)
% increase in risk calculated as (1-hazard ratio)x100%; separate model for each cancer with linear BMI term, all models adjusted for adjusted for age, diabetes status, smoking, alcohol, socioeconomic status, calendar year, and stratified by sex.

Figure W1.9: Estimated linear BMI-cancer association (HR per 5kg/m2 increase in BMI and 99% CI) with follow-up increasingly restricted to more recent calendar periods with higher BMI completeness



% increase in risk calculated as (1-hazard ratio)x100%; separate model for each cancer with linear BMI term, all models adjusted for adjusted for age, diabetes status, smoking, alcohol, socioeconomic status, calendar year, and stratified by sex.
"ALL" = original estimate using all follow-up in CPRD; "1995-" restricts to follow-up from 1995 onwards (events and person-time before 1st January 1995 were excluded"; similarly for "2000-" and "2005-" in which

Figure W1.10: Post-hoc sensitivity analysis censoring at first use of hormone replacement therapy, for models relating BMI to postmenopausal breast and ovarian cancer



Webappendix Part 2 – Original study protocol

Notes:

- Protocol submitted to the Medicines and Healthcare products Regulatory Agency (MHRA) Independent Scientific Advisory Committee (ISAC) – protocol number 12_090, approved 30 July 2012 (approval appended)
- The present manuscript covers the material described in "Aim 2"; the material in "Aim 1" was written up as a separate manuscript (Bhaskaran et al, BMJ Open 2013)
- A list of deviations from the protocol with explanations can be found on the last page of this section. A formal protocol amendment was approved in October 2013 to cover changes to the original study plans.

Protocol: Body mass index data in CPRD, and the association between body mass index and specific cancers

Krishnan Bhaskaran *(lecturer in statistical epidemiology)*
Ian Douglas *(lecturer in pharmacoepidemiology)*
Harriet Forbes *(research assistant in epidemiology)*
Liam Smeeth *(professor of clinical epidemiology)*

Department of non-communicable diseases epidemiology, London School of Hygiene and Tropical Medicine, London, UK

Lay summary
Body mass index (BMI) is the measure most often used by doctors to quantify individual body fatness, and a number of studies have suggested that higher BMI is associated with an increased risk of certain cancers. BMI is often routinely recorded by general practitioners, and by linking BMI records with later medical records for the large number of patients included in the Central Practice Research Datalink (CPRD), we plan to quantify the association between BMI and a wide range of specific cancers. The volume of data available within CPRD will enable us to characterise the relationships between BMI and different cancers accurately, and to investigate whether these relationships differ depending on individual characteristics such as age, gender, smoking and alcohol status. In order to inform our main analysis and understand the limitations of the data available, we will first carry out a preliminary analysis looking at the proportion of people with BMI data available in CPRD and their characteristics; then by comparing BMI data in CPRD with published BMI data from population-based surveys data, we will assess to what extent those with BMI data recorded in CPRD are likely to be representative of the wider population.

Aims and objectives
Aims

1) To explore the completeness and validity of BMI data in CPRD
2) To use the large amount of BMI and clinical data available in CPRD, along with linked data sources, to describe and quantify the associations between BMI and specific cancers.

Specific objectives:

1) To describe the completeness of BMI data in CPRD over calendar time, within age and sex strata, and overall
2) To explore the validity/representativeness of BMI data in CPRD by comparing summary BMI statistics based on non-missing BMI data in CPRD (English practices) with summary statistics published by the Health Survey for England
3) To quantify the association between BMI and cancer, for a wide range of cancer types
4) To explore modification of BMI-cancer associations by individual-level factors

Background and rationale

The increasing prevalence of overweight and obesity is a growing concern for policy-makers both in the UK and globally. Recent data from England showed the prevalence of obesity rising steadily from 1993 to 2009: from 13% to 22% in men, and from 16% to 24% in women.[1] In 2009, an estimated 61% of adults in England were overweight or obese. Overweight and obesity are also rising globally.[2]

BMI (=weight/height^2) is the metric most widely used to quantify adiposity. The importance of BMI as a predictor of various adverse health outcomes is recognised; in particular associations between BMI and a number of cancers have been reported. However, the picture to date has been built up from a series of mostly small and heterogeneous studies. Several meta-analyses collating the evidence for site-specific cancers have suggested associations between increasing BMI and risk of colorectal,[3] liver,[4] gallbladder,[5] and gastro-oesophageal cancers,[6, 7] as well as Hodgkin and non-Hodgkin lymphomas.[7] A few noteable large individual studies have also published data, including a cohort study based on Norwegian survey data that reported associations between BMI and a number of cancers, including thyroid[8] and ovarian;[9] and an analysis using data from the Million Women Study which found an association between BMI and 10 out of 17 specific cancers among women aged 50-64 at study entry.[10] A 2007 review of the evidence by the World Cancer Research Fund,[11] and a later systematic review published in the Lancet,[12] brought much of these disparate data together. In a meta-analysis including 221 datasets focussing on BMI and cancer outcomes, the latter study reported that the strongest associations were found between BMI and cancers of the oesophagus, thyroid, colon, kidneys, endometrium, and gallbladder, and weaker associations were found for a number of other malignancies. Increased BMI was negatively associated with lung cancer. However, the review's authors noted a number of limitations: studies using self-reported BMI data (likely to underestimate true BMI[13]) were included; there was some evidence of differences between the sexes in BMI/cancer relationships, but there was insufficient information from the original studies to determine the role of other factors such as age and smoking status; and confounding by factors that were inconsistently measured across studies (such as smoking) or that may not have been measured at all (such as hormone replacement therapy use, in the context of breast cancer), could not be excluded.

We plan to investigate the links between BMI and specific cancers within CPRD. BMI is often recorded as part of routine general practice, making CPRD a very large source of BMI data, with relevant cancer outcomes data available both within the

database itself and in linked data sources including the cancer registries, and comprehensive data on a wide range of potential confounders including lifestyle factors, treatments, and co-morbidities. To date these data have not been widely used to contribute to the evidence base surrounding BMI-cancer associations. A key advantage is that the volume of data available will allow us to quantify associations with considerable precision, and to investigate a wide range of specific cancers, including relatively uncommon malignancies, within a unified methodological framework, facilitating comparison of the relative importance of BMI to risks of specific cancers. We will also be able to explore the role of individual-level effect modifiers including age, smoking and co-morbidities, which has not been satisfactorily possible with the data available to date, as noted by Renehan et al in their wide-ranging review.[12]

An important first step will be to investigate the completeness of BMI data in CPRD, and to what extent those with BMI data available are likely to be broadly representative of the general population. We therefore plan to carry out a preliminary study (Aim 1) quantifying the completeness of BMI data in CPRD, and comparing estimates of population BMI summaries based on CPRD with those published by the Health Survey for England, an annual survey designed to be representative of the general population. Our findings from this preliminary stage will allow us to optimise our main analysis looking at BMI-cancer associations, and will enable us to better understand and discuss the limitations of the data available.

Study type: Hypothesis generating

Aim 1. Completeness and validity of BMI data in CPRD

NB THIS PART OF THE PROTOCOL (AIM 1) REFERS TO ANALYSES THAT ARE PUBLISHED SEPARATELY (Bhaskaran et al, BMJ Open 2013)

Study design

Cross-sectional descriptive analysis at multiple calendar time points, based on a large random sample of CPRD data (see "Sample size" below).

Variables of interest and data processing

The main variable of interest is BMI. The vast majority of BMI data in CPRD are system-generated from raw height and weight measurements. For the purposes of this study, we will obtain the original weight and height records and re-calculate BMI, to ensure consistent and reproducible data management, for example in dealing with height or weight measurements that are implausible or missing. We will obtain all height and weight records from age 16 years and over. BMI will be calculated directly where height and weight are recorded on the same day (using BMI = weight/height^2). For the remainder of weight records, the most recent past height measurement will be used, or if not available, the first future height measurement.

Data/Statistical analysis

Completeness of BMI data

BMI completeness data in CPRD will be generated by calendar period (1990-4, 1995-9, 2000-4, 2005-11), and by individual calendar year. To calculate completeness data for a particular calendar period, all individuals in the sample who are registered, aged ≥16 years, and under follow-up in up to standard practices on the date of the mid-point of the period will be identified and included in the denominator. Among these individuals, the number with any BMI available on or prior to this date, and the number with a BMI available up to 3 years prior to this date will be counted in order to calculate completeness with any previous BMI and with a recent BMI. Completeness data will be generated by age group (16-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+ years), by sex, and among those with a record of specific clinical conditions in which BMI is expected to be monitored (namely type 2 diabetes, schizophrenia/other psychoses, bipolar disorder). The proportions of newly GP-registered patients having a BMI entered at 3, 6, 9, 12, and 15 months following GP registration will be estimated, stratified by calendar period of registration, using a Kaplan-Meier-based cumulative incidence function (with the "failure" event being having a BMI entered in the database).

Comparison of CPRD BMI data with Health Survey for England (HSE) data

We will compare mean BMI over calendar time based on complete CPRD BMI data with equivalent figures published in the HSE tables, for the period 2003-2010 (for these calendar years, HSE data adjusted for non-response are available). For each calendar year, the CPRD mean BMI will be based on patients registered and under up-to-standard follow-up at the mid-point of the calendar year. Two sets of CPRD mean BMI statistics will be generated: the first using the most recent BMI for all patients with a previous BMI available, regardless of how long ago it was entered; the second restricted to patients with a recent BMI available (up to 3 years before the mid-point of the calendar year). To aid comparison with HSE, CPRD data will be restricted to English practices, and mean BMI will be age- and sex-standardised to the HSE population structure. Proportions classified as obese (BMI≥30kg/m^2) over time based on CPRD and HSE data will also be compared (similarly age- and sex- standardised to the HSE population structure). Unstandardised stratum-specific estimates will also be compared.

Correction of outdated BMI records

We will explore whether outdated BMI measures in CPRD can be usefully updated based on a model predicting changes in individual-level BMI over time. We will fit a linear mixed model using all available BMI measurements (including multiple measurements per patient), including age (with random effects at the patient level), sex, and calendar year. We will then

"correct" individual outdated BMI measures based on the model-based prediction for the age- and calendar year- associated change in BMI. We will explore the performance of this correction by first repeating the comparison with the HSE data, using the time-corrected CPRD data; and second by using the model-based correction to predict the most recent BMI using the previously-entered BMI, among patients with at least 2 BMIs entered in CPRD; we will summarise the distribution of the error arising from the model-based corrected BMI, and compare with the error distribution obtained from simply carrying the older BMI forward without correction.

Sample size
We will obtain a sample based on one million randomly selected CPRD patient identifiers. This makes computation more practical, compared with using the full database as the denominator, and the sampling error associated with a sample of this size is negligible for our purposes. For example, initial data exploration indicate that after stratification by calendar period and age group, denominators for the completeness data will include over 10000 individuals in the smallest strata, making the associated confidence interval widths for proportions with complete data no bigger than 0.0062 [$= 2x1.96x\sqrt{(0.5x0.5/10000)}$], i.e. 0.62%). For the comparison of CPRD-based and HSE-based summary statistics, the smallest base population in a calendar year will be 89056, and assuming a standard deviation for BMI of $<6 kg/m^2$, which is consistent with the data, the confidence interval width for mean BMI will be $<0.08 kg/m^2$ [$= 2x1.96x6/\sqrt{89056}$].

Limitations of the study design, data sources and analytic methods
A limitation of using Health Survey for England data as the comparator to assess validity is that these data are only representative of those living in private households; individuals living in institutions are not represented. Therefore the data for the very elderly, many of whom may be living in institutions, may not be comparable. We will acknowledge and discuss this limitation in any manuscript.

<u>Aim 2. Quantifying the associations between BMI and specific cancers</u>

Study design
Cohort study using CPRD data and linked HES, cancer registry and ONS mortality data.

Primary exposure
The main exposure variable will be BMI at start of follow-up (see Statistical analysis for details on assignment of exposure status). We will calculate BMI based on raw weight and height measurements, following the algorithm outlined in "Aim 1 - Variables of interest and data processing".

Outcomes
The outcomes of interest will be first occurrence of specific cancers (based on the 20 most commonly diagnosed cancers in the UK[14] and cancers with previous evidence of association with BMI[12]): breast (ICD code = C50), prostate (C61), colorectal (C18-20), lung (C34), malignant melanoma (C43), bladder (C67), gastric (C16) and oesophageal (C16), non-Hodgkin lymphoma (NHL, C82-85), leukaemia (C91-95), ovary (C56), pancreas (C25), myeloma (C90), uterus (C54-55), brain/central nervous system (brain/CNS, C71-72), liver (C22), kidney (C64), cervix (C53), oral (C00-06), thyroid (C73) and gallbladder(C23). As part of previous work,[15] we have identified medical codes from the CPRD medical dictionary related to cancer, and mapped these to ICD-10 chapter 2 headings. This mapping will be used to identify cases of the above cancers in CPRD based on ICD code. (See also "Sensitivity analysis using alternative data sources for case ascertainment" below).

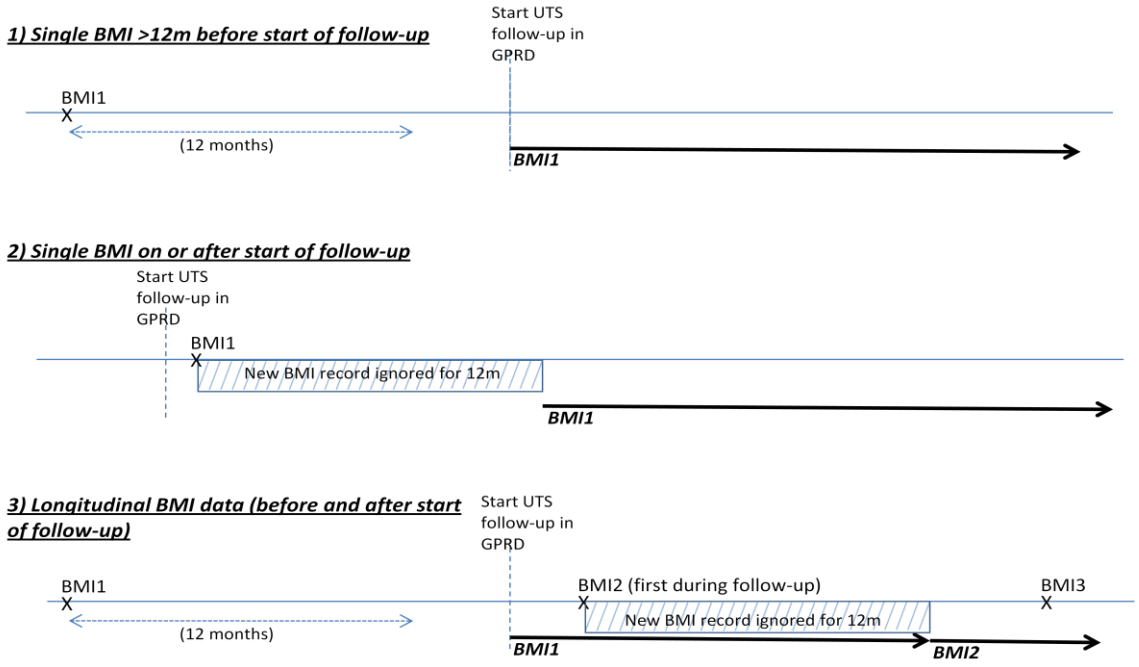
Statistical analysis
We will conduct a time-to-event analysis with attained age as the underlying timescale. For each patient, the risk period will begin at the latest of: current registration date, practice up to standard date, sixteenth birthday, date of first recorded BMI in CPRD; and follow-up will end at the earliest of: last collection date for practice, occurrence of any cancer, death or transfer out of CPRD.

To assign exposure, we will use the first BMI recorded in CPRD. The 12 month period following the date of this BMI record will be excluded from the risk period to guard against reverse causality (i.e. the use of BMI measures that might have been affected by undiagnosed cancers, Figure 1). If the first BMI recorded in CPRD is before the start of up-to-standard CPRD follow-up, then this BMI will be assigned to the patient from the date of up-to-standard CPRD follow-up* (or 12 months after the BMI measure, if later), and if a further BMI record exists during up-to-standard follow-up, we will time-update the exposure assignment 12 months after the date of the first BMI during up-to-standard follow-up. Second and subsequent BMI records during up-to-standard follow-up will not be used, other than in a sensitivity analysis (see "Sensitivity analysis using all available longitudinal BMI measures" below).

* If our preliminary work (Aim 1) indicates that a model-based correction of outdated BMI measures would provide a better estimate of current BMI than simply carrying the last measure forward, then we will employ such a correction to assign current BMI at start of follow-up. Any such approach, if taken, will be compared in a sensitivity analysis with a standard approach of carrying the last measure forward.

We will calculate crude incidence rates for each cancer type in strata of BMI, and we will then fit Cox and Poisson models to estimate relative and absolute risks, adjusting for a wide range of potential confounders, tailored to specific cancer types but including sex, smoking status, alcohol use, socioeconomic status, calendar period, age at BMI recording, and relevant co-morbidities and medication use. Effect modification will be explored by fitting appropriate interaction terms; of specific a priori interest will be effect modification by age, sex and smoking status; we will also explore the role of comorbidities including diabetes. As well as summarising the evidence for a simple dose-response relationship between BMI and risk through assuming linear relationships, we will also investigate possible non-linearity by fitting flexible splines, and other flexible semi-parametric models,[16] as well as BMI categories. Attributable fractions and population attributable fractions will be calculated based on BMI categories.

Figure 1: Assignment of exposure in 3 example scenarios



Notes: Exposure assignment is shown by the bold arrows.
Scenario (1) – patient enters the risk set at start of UTS follow-up because there is a BMI measure available (BMI1) and more than 12m have elapsed since it was recorded. Their exposure is assigned as BMI1 throughout.
Scenario (2) - patient has a BMI recorded after start of UTS follow-up, but we will only enter them into the risk set 12m after the BMI recording, to guard against reverse causality. Exposure is then assigned as BMI1 throughout the rest of follow-up.
Scenario (3) - patient enters the risk set at start of UTS follow-up because there is a BMI measure available (BMI1) and more than 12m have elapsed since it was recorded. Exposure is assigned as BMI1 at start of follow-up (as per scenario 1). During follow-up a new BMI is recorded (BMI2): the exposure remains BMI1 until 12m after the new BMI measure, at which point the exposure updates to BMI2. A third BMI recording (BMI3) and any further BMIs are not used.

Further sensitivity and exploratory analyses

Sensitivity analysis including those with no BMI available
The exclusion of individuals with no BMI available in CPRD may lead to selection bias. For example, those without a BMI recorded may be more likely to have a healthy BMI that has not concerned their GP; or, these may be individuals who tend to visit their GP less. To explore this, we will carry out a sensitivity analysis in which we include these individuals, assigning their BMI as missing. We will then impute BMI measures using a multiple imputation model including possible predictors of BMI and BMI missingness (including age, sex, smoking status, alcohol status, socioeconomic status, calendar year, morbidities, GP consultation rate).

Sensitivity analysis only using BMIs in first 12m after registration
A proportion of patients in our main analysis will have had their first BMI record entered some time after their current GP registration. It is possible that such BMI recordings were made for health reasons (e.g. the GP became concerned about the person's weight). Our policy of excluding the 12 months after the first BMI from the risk set will reduce the danger of reverse causality, but as a further check that important selection biases are not being introduced, we will carry out a sensitivity analysis restricted to patients whose first BMI was recorded within 12 months of their current registration, as BMIs recorded at or soon after registration are more likely to have been recorded for administrative rather than health reasons.

Sensitivity analysis using all available longitudinal BMI measures
In our main analysis, we will only use the first BMI measure during follow-up, because individuals with multiple longitudinal BMI measures may be atypical. However, in a sensitivity/exploratory analysis, we will time-update exposure status using all available longitudinal information.

Sensitivity analysis using alternative data sources for case ascertainment
In a further set of sensitivity analyses restricted to practices with linked data available, we will use HES, cancer registry, and ONS to identify cancers, and explore whether the use of these alternative data sources to identify cancer cases changes our main conclusions.

Exploratory analysis using data from the first 12m after BMI recordings
In the main analysis, we are excluding the 12 months after a patients first BMI record, in case of reverse causality (on the basis that cancers diagnosed soon after a BMI record may have affected BMI at the pre-diagnosis stage). In a separate exploratory analysis we will explore the extent of likely reverse causality and bias by specifically looking at the association between BMI and risk of cancer in the 12 months after the BMI record (i.e. the period excluded from the main analysis).

Sample size/power calculation
We plan to use all available data in order to estimate and quantify the BMI-cancer association for each cancer type with maximum precision. Preliminary event counts have been carried out in a one million random sample of CPRD and are extrapolated in Table 1 to summarise the expected counts in the full database; these counts refer to cancer events occurring during up-to-standard follow-up and with a valid BMI based on the risk period and BMI date restrictions outlined in the Statistical Analysis section above.

Also shown in Table 1 are the estimated minimum detectable effect sizes for each cancer type based on a Cox regression analysis, using the method of Hsieh and Lavori (implemented in Stata's "stpower cox" command).[17] Based on preliminary data, we estimate the standard deviation of BMI to be 5.7kg/m^2, which is incorporated in the power calculations. The number of events available will allow us to detect with formal statistical significance a hazard ratio (HR) of 1.15 per 5kg/m^2 increase in BMI for the least common cancer included (gallbladder), and minimum hazard ratios between 1.012 and 1.078 per 5kg/m^2 increase in BMI for the remaining individual cancers.

These figures demonstrate that we will be able to estimate the associations between BMI and cancer with good precision for the majority of specific cancers, enabling us to quantify most BMI-cancer associations with greater statistical precision than has yet been achieved within an individual study/data source.

Table 1: Estimated event counts for specific cancers in the full CPRD database

Cancer site/type	ICD codes	N events during follow-up and with BMI available*	Min detectable HR (per 5kg/m2 increase in BMI)**
Breast	C50	40090	1.012
Prostate	C61	27968	1.015
Colorectal	C18-20	22782	1.017
Lung	C34	22713	1.017
Malignant Melanoma	C43	9315	1.026
Bladder	C67	9860	1.025
Non-Hodgkin Lymphoma	C82-85	8886	1.027
Leukaemia	C91-95	7969	1.028
Oesophageal	C15	6020	1.032
Ovary	C56	4733	1.037
Pancreas	C25	4199	1.039
Gastric	C16	4199	1.039
Myeloma	C90	3677	1.042
Uterus	C54-55	3700	1.042
Brain/CNS	C71-72	3364	1.044
Liver	C22	2297	1.053
Kidney	C64	2274	1.053
Cervix	C53	1775	1.061
Oral	C00-06	1485	1.066
Thyroid	C73	1090	1.078
Gallbladder	C23	325	1.147

*These are the estimated number of events in the full CPRD based on counts taken from a one million sample of CPRD
**Based on 80% power, 0.05 type I error rate, sd(BMI) = 5.7kg/m^2

Limitations of the study design, data sources and analytic methods

A potential limitation of our planned study investigating the associations between BMI and cancers is that, unless BMI data are missing completely at random, we could introduce a selection bias through excluding those without a BMI recording. Through Aim 1 of this protocol we will have comprehensive information on the completeness of BMI data, and how representative those with BMI recorded appear to be of the wider population; we will thus be in a strong position to acknowledge and discuss the likely biases, and if appropriate, to develop strategies for reducing them. We have outlined one such strategy above, namely to include those with missing BMI data in a sensitivity analysis, in which we will impute the missing BMI data within a multiple imputation framework. In this analysis we will have greater confidence that the overall study population is representative of the general population (since we believe CPRD as a whole to be broadly representative), while the confidence intervals around our effect estimates will appropriately reflect the missing BMI data.

A further limitation is that cancer cases may be missed, ambiguously coded, or ambiguously dated in the CPRD record. As outlined above, we therefore plan to repeat our analyses among patients in linked practices and using cancer registry, HES and ONS data for case ascertainment, and to check that our overall conclusions can be confirmed.

Patient/user group involvement

We do not believe this research would benefit from patient group involvement at this stage, although we will actively collaborate with such groups in the dissemination strategy. It is possible that future research may well benefit from such involvement.

Plans for disseminating and communicating study results

Clinical findings will be disseminated through presentation at international conferences, and through publication in the relevant medical journals. We also plan to publish our findings on BMI data completeness and validation in a relevant epidemiology journal.

References

1. NHS Information Centre. Health survey for England - 2009: health and lifestyles. 2010 [cited 2011 1st February]; Available from: www.ic.nhs.uk/pubs/hse09report
2. World Health Organization. World Health Statistics 2011.
3. Larsson SC, Wolk A. Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies. Am J Clin Nutr. 2007; **86**(3): 556-65.
4. Larsson SC, Wolk A. Overweight, obesity and risk of liver cancer: a meta-analysis of cohort studies. Br J Cancer. 2007; **97**(7): 1005-8.
5. Larsson SC, Wolk A. Obesity and the risk of gallbladder cancer: a meta-analysis. Br J Cancer. 2007; **96**(9): 1457-61.
6. Kubo A, Corley DA. Body mass index and adenocarcinomas of the esophagus or gastric cardia: a systematic review and meta-analysis. Cancer Epidemiol Biomarkers Prev. 2006; **15**(5): 872-8.
7. Larsson SC, Wolk A. Body mass index and risk of non-Hodgkin's and Hodgkin's lymphoma: a meta-analysis of prospective studies. Eur J Cancer. 2011; **47**(16): 2422-30.
8. 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**(3): 366-70.
9. 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**(16): 1244-8.
10. 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**(7630): 1134.
11. 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.
12. 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**(9612): 569-78.
13. 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**(4): 561-5.
14. Cancer Research UK. Cancer incidence for common cancers. 2012 [cited 2012 20th June]; Available from: http://info.cancerresearchuk.org/cancerstats/incidence/commoncancers/
15. Bhaskaran K, Douglas I, Evans S, van Staa T, Smeeth L. Angiotensin receptor blockers and risk of cancer: cohort study among people receiving antihypertensive drugs in UK General Practice Research Database. BMJ. 2012; **344**: e2697.
16. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med. 2002; **21**(15): 2175-97.
17. Hsieh FY, Lavori PW. Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. Controlled Clinical Trials. 2000; **21**(6): 552-60.

ISAC EVALUATION OF PROTOCOLS FOR RESEARCH INVOLVING GPRD DATA

FEED-BACK TO APPLICANTS

CONFIDENTIAL			***by e-mail***
PROTOCOL NO:	12_090		
PROTOCOL TITLE:	Body mass index data in CPRD, and the association between body mass index and specific cancers		
APPLICANT:	Dr Krishnan Bhaskaran, Lecturer in statistical epidemiology, London School of Hygiene and Tropical Medicine, krishnan.bhaskaran@lshtm.ac.uk		
APPROVED ☒	**APPROVED SUBJECT TO MINOR AMENDMENT (resubmission not required)** ☐	**REVISION/ RESUBMISSION REQUESTED** ☐	**REJECTED** ☐

INSTRUCTIONS:

Please include your response/s to the Reviewer's feedback below only if you are required to Revise/ Resubmit your protocol.

Protocols with an outcome of 'Approved' or 'Approved subject to minor amendments' do not require resubmission to the ISAC

REVIEWER COMMENTS:

No amendment to the protocol is necessary but you may wish to take on board the following comments:

1. BMI data may also be recorded using Read codes so in instances where weight and height data or BMI values are missing you may wish to consider the use of Read coded information.

2. Where values are implausible it may be helpful to know that GPs can record weight and height information in Vision using either imperial measurement or SI units. The system default is to treat all measurements as SI unless the GP declares the units to be otherwise. This may assist with the 'correction' of implausible values.

3. In defining the start and end of follow-up of patients for the cancer study, you should incorporate the start and end date of the linked data sources in the equation.

DATE OF ISAC FEEDBACK:	30 July 2012
DATE OF APPLICANT FEEDBACK:	

SUMMARY OF DEVIATIONS FROM THE PROTOCOL AND JUSTIFICATION

(1) to (5) below were approved as a formal protocol amendment (protocol number 12_090A) by the Independent Scientific Advisory Committee (ISAC) on 28th October 2013. It was agreed with the chair of ISAC that item (6) did not require a formal amendment.

1) We decided to exclude the first 12 months of follow-up time in CPRD completely because prevalent cancer cases might be recorded in the database shortly after GP registration, and misclassified as incident events.

2) We decided to separate colorectal cancers into colon (C18) and rectum (C20) because of evidence in the literature of differences between them in terms of associations with BMI.

3) For cancers of the uterus, cervix and ovaries, we decided to censor at hysterectomy since patients would not be at risk of these cancers after this point (we assumed all hysterectomies to involve removal of the cervix and ovaries). We allowed 30 days after the first record of a hysterectomy since cancers diagnosed just before the hysterectomy (or detected at the time of hysterectomy) might be entered in the database just after, and would otherwise be excluded.

4) We used Hospital Episodes Statistics and Death Certificate data as an additional way of detecting cancer outcomes in a sensitivity analysis, as planned, but we decided not to additionally use UK Cancer Registry data as mentioned in the protocol. This is because our previous use of the cancer registry database revealed that these data were some years out of date, which would have resulted in much reduced power. Recent data published by Boggon et al (Pharmacoepidemiology and Drug Safety 2013) also suggested that 94% of cancers recorded in national cancer registries were recorded in CPRD and over 99% were in CPRD or hospital/death certificate data, suggesting that omitting this data source would be expected to have very little impact.

5) The proposed multiple imputation analysis involved imputation for around 4 million individuals and presented computing problems, so we did not include it. Its principal motivation had been to address selection bias by enabling inclusion of patients without a BMI. Instead we addressed selection bias by (i) restricting to patients with a BMI recorded within 12 months of GP registration (as mentioned in the protocol); (ii) adding a sensitivity analysis with adjustment for GP consultation rate, to block proposed biasing paths resulting from restriction to those with a BMI measure. We have also extensively explored the representativeness of those with BMI recorded in CPRD by comparing with representative survey data as part of Aim 1 of this protocol (published as Bhaskaran et al, BMJ Open 2013, In Press); we have cited and discussed this in the manuscript.

6) It was decided on advice from a co-author (IdSS) to split female breast cancer *a priori* into pre-menopausal and post-menopausal, because there was already strong evidence in the literature of a difference in the effect of BMI on each outcome.

Webappendix Part 3 – Systematic literature review for Discussion
Flow chart of review process, detailed table of studies, reference list

Figure W3.1: Flow chart of systematic review search and included studies

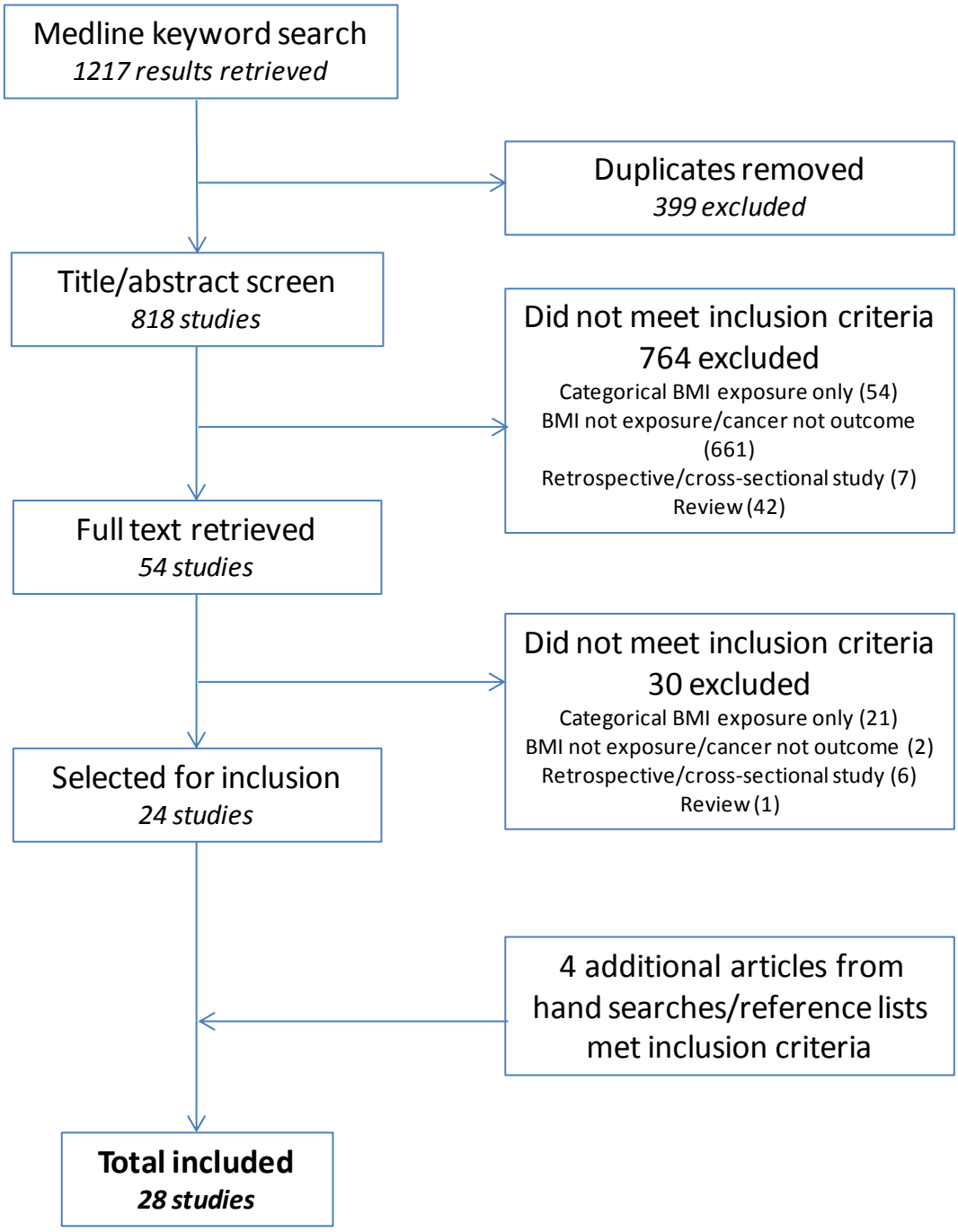


Table W3.1: Comparison of estimated BMI-cancer associations with those from previous[1] and updated systematic literature review

CANCER TYPE (ICD10) 1st Author & Year	Design	Location	N (mean yrs follow-up)	Overall RR per 5kg/m^2 BMI increase (CI)	Effect modification and other comments (RRs where quoted are scaled to per 5kg/m^2 increase in BMI)
OESOPHAGEAL (C15)					
Present study	Cohort	UK	5.2 million (9)	1.03 (0.99, 1.08)*	U-shaped BMI-cancer association: inverse association at low BMIs; positive association at higher BMIs, especially in men, non-smokers, and those aged <75y
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	10 studies	Not reported	Adenocarcinoma - Men: 1.52 (1.33, 1.74); Women: 1.51 (1.31, 1.74) Squamous – Men: 0.71 (0.60, 0.85); Women: 0.57 (0.47, 0.69)
STOMACH (C16)					
Present study	Cohort	UK	5.2 million (9)	1.03 (0.98, 1.09)*	Inverse association at low BMIs, attenuated in the oldest age group
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	10 studies	Not reported	Men: 0.97 (0.88, 1.06); Women: 1.04 (0.90, 1.20)
Song 2008[2]	Cohort	Korea	170,481 (9)	N/A	Post-menopausal women only: 1.00 (0.95, 1.10)**
COLON (C18)					
Present study	Cohort	UK	5.2 million (9)	1.10 (1.07, 1.13)*	Men: RR=1.23 (1.17-1.30)* per 5kg/m^2 between 22 and 34kg/m^2 only Women: 1.05 (1.02-1.08)* per 5kg/m^2
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	29 studies	Not reported	Men: 1.24 (1.20, 1.28); Women: 1.09 (1.05, 1.13)
Song 2008[2]	Cohort	Korea	170,481 (9)	N/A	Post-menopausal women only: 1.28 (1.10, 1.47)
Harris 2009[3]	Meta-analysis	Mostly Europe/US	28 studies	Not reported	Men: 1.24 (1.20, 1.28); Women: 1.09 (1.05, 1.14)
Laake 2010[4]	Cohort	Norway	76,179 (23)	Not reported	Colon proximal – Men: 1.07 (0.86, 1.33); Women: 1.15 (0.99, 1.34) Colon distal – Men: 1.49 (1.19, 1.87); Women: 1.25 (1.05, 1.49) In women, significant effect post-menopausal only
Bassett 2010[5]	Cohort	Australia	39,626 (26)	Not reported	(Colon only) – Men: 1.39 (1.12, 1.71); Women (1.04 (0.98, 1.11)
Hughes 2011[6]	Cohort	Netherlands	120,852 (16)	Not reported	Proximal colon – Men: 1.19 (0.92, 1.54); Women (1.02 (0.87, 1.18) Distal colon – Men: 1.42 (1.13, 1.79); Women 0.95 (0.79, 1.14)
Renehan 2012[7]	Cohort	US	273,679 (9)	Not reported	Colon – Men: 1.18 (1.11, 1.25); Women: 1.05 (0.98, 1.12) In women, risk associated with BMI among current HRT users (RR 1.13, 1.01-1.26) but not non-users
Semmens 2013[8]	Cohort	Japan	56,064	1.14 (1.03, 1.26)	Men: 1.25 (1.07, 1.45); Women: 1.07 (0.94, 1.22); study of atomic bomb survivors
RECTUM (C20)					
Present study	Cohort	UK	5.2 million (9)	1.04 (0.97, 1.08)*	No evidence of non-linearity or effect modification
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	29 studies	Not reported	Men: 1.09 (1.06, 1.12); Women: 1.02 (1.00, 1.05)
Song 2008[2]	Cohort	Korea	170,481 (9)	N/A	Post-menopausal women only: 1.00 (0.86, 1.16)**
Harris 2009[3]	Meta-analysis	Mostly Europe/US	28 studies	Not reported	Men: 1.09 (1.05, 1.14); Women: 1.02 (0.99, 1.04)
Hughes 2011[6]	Cohort	Netherlands	120,852 (16)	Not reported	Men: 1.02 (0.79, 1.32); Women 1.05 (0.83, 1.31)
Renehan 2012[7]	Cohort	US	273,679 (9)	Not reported	Men: 1.03 (0.93, 1.14); Women: 1.05 (0.92, 1.19)

LIVER (C22)

Present study	Cohort	UK	5.2 million (9)	1.19 (1.12, 1.27)*	Men: BMI<22kg/m^2: RR 0.57 (0.33, 1.00)* BMI≥22kg/m^2: RR 1.31 (1.19, 1.43)* Women: RR 1.15 (1.04, 1.26)
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	5 studies	Not reported	Men: 1.24 (0.95, 1.62); Women: 1.07 (1.00, 1.14)
Song 2008[2]	Cohort	Korea	170,481 (9)	N/A	Post-menopausal women only: 1.05 (0.90, 1.16)**
Tanaka 2012[9]	Meta-analysis	Japan	9 studies, N=160,633	1.40 (1.16, 1.61)	(Note, review was restricted to studies of Japanese populations)

GALLBLADDER (C23)

Present study	Cohort	UK	5.2 million (9)	1.31 (1.12, 1.52)*	No evidence of non-linearity or effect modification
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	5 studies	Not reported	Men: 1.09 (0.99, 1.21); Women: 1.59 (1.02, 2.47)
Song 2008[2]	Korea	Cohort	170,481 (9)	N/A	Post-menopausal women only: 1.22 (0.95, 1.61)**

PANCREAS (C25)

Present study	Cohort	UK	5.2 million (9)	1.05 (1.00, 1.10)*	No evidence of non-linearity or effect modification
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	16 studies	Not reported	Men: 1.07 (0.93, 1.23); Women: 1.12 (1.02, 1.22)
Song 2008[2]	Cohort	Korea	170,481 (9)	N/A	Post-menopausal women only: 1.00 (0.82, 1.28)**
Aune 2012[10]	Meta-analysis	Mostly Europe and US	23 studies N=5.0 million	1.10 (1.07, 1.14)	Non-linear association reported (lowest risk around BMI = 21kg/m^2, most pronounced risk increase above BMI=35kg/m^2); but some differences by smoking status

LUNG (C34)

Present study	Cohort	UK	5.2 million (9)	0.82 (0.80., 0.84)*	Inverse BMI-cancer associations at lower BMIs primarily driven by current and ex-smokers.
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	13 studies	Not reported	Men: 0.76 (0.70, 0.83); Women: 0.80 (0.66, 0.97)
Song 2008[2]	Cohort	Korea	170,481 (9)	N/A	Post-menopausal women only: 0.86 (0.73, 0.95)**

MALIGNANT MELANOMA (C43)

Present study	Cohort	UK	5.2 million (9)	0.99 (0.96., 1.02)*	Men: BMI<24kg/m^2: RR 1.48 (1.17, 1.87)* BMI≥24kg/m^2: RR 0.99 (0.93, 1.06)* Women: RR 0.96 (0.92, 1.00)
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	7 studies	Not reported	Men: 1.17 (1.05, 1.30); Women: 0.96 (0.92, 1.01)

BREAST (C50)

Present study	Cohort	UK	5.2 million (9)	*Pre-menopause:* 0.89 (0.86, 0.92)* *Post-menopause:* 0.95 (0.93, 0.97)*	Pre-menopause: BMI<22kg/m^2: RR 1.20 (1.01, 1.43)* BMI≥22kg/m^2: RR 0.86 (0.82, 0.89)* Post-menopause: BMI<29kg/m^2: RR 1.11 (1.08, 1.14)* BMI≥29kg/m^2: RR 0.98 (0.95, 1.01)*
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	34 studies	Not reported	Pre-menopause: 0.92 (0.88-0.97); post-menopause: 1.12 (1.08, 1.16)
Song 2008[2]	Cohort	Korea	170,481 (9)	N/A	Post-menopausal women only: RR 1.40 (1.28 1.61)**
Suzuki 2009[11]	Meta-analysis	Mainly Europe/N America	23 studies	Not reported	For ER+PR+ tumours, pre-menopause RR 0.90 (0.82, 0.99) post-menopause 1.33 (1.20, 1.48); for ER-PR- ER+PR- no effect
Bjorge 2010[12]	Cohort	Austria, Norway, Sweden	287,320 (11)	Not reported	Age<50: RR 0.84 (0.78, 0.91); Age 50-59 0.96 (0.89, 1.02); Age 60+ 1.08 (1.02, 1.15)**
Ritte 2012[13]	Western Europe	Cohort	314,676 (11)	Not reported	For ER+PR+ tumours, RR ranged from 0.80 (0.69, 0.93) among those <50 to 1.32 (1.22, 1.43) among those 65+; for ER-PR-, no clear association with BMI

CERVIX (C53)					
Present study	Cohort	UK	5.2 million (9)	1.10 (1.03, 1.17)*	No evidence of non-linearity or effect modification
Song 2008[2]	Cohort	Korea	170,481 (9)	N/A	Post-menopausal women only: RR 1.10 (0.95, 1.28)**
Ulmer 2012[14]	Cohort	Austria, Norway, Sweden	288,834 (11)	RR 1.14 (95% CI 1.01-1.29)	-
Reeves 2007[15]	Cohort	UK	1.2 million (5)		Women aged >50 only: RR 1.02 (0.89, 1.16)**
UTERUS (C54-55)					
Present study	Cohort	UK	5.2 million (9)	1.62 (1.56, 1.69)*	No evidence of non-linearity or effect modification
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	19 studies	1.59 (1.50, 1.68)	-
Song 2008[2]	Cohort	Korea	170,481 (9)	N/A	Post-menopausal women only: 1.84 (1.40, 2.49)**
Lindemann 2008[16]	Cohort	Norway			
Crosbie 2010[17]	Meta-analysis	Mainly Europe/ N America	24 studies	1.60 (1.52, 1.68)	Never HRT: 1.90 (1.57–2.31); Ever HRT 1.18 (1.06–1.31). No effect modification by menopausal status.
Yang 2012[18]	Cohort	United Kingdom	249,791 (7)	1.87 (1.77, 1.96)	No evidence of effect modification by early body size
OVARIES (C56)					
Present study	Cohort	UK	5.2 million (9)	1.11 (1.06, 1.16)*	Pre-menopause: 1.21 (1.04, 1.26)* Post-menopause: 1.07 (1.02, 1.12)*
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	13 studies	1.03 (0.99, 1.08)	-
Song 2008[2]	Cohort	Korea	170,481 (9)	N/A	Post-menopausal women only: 1.22 (0.95, 1.54)**
Chionh 2010[19]	Cohort	Australia	18,700 (10)	1.22 (1.00, 1.48)	No evidence of effect modification by physical activity
Beral 2012[20]	Meta-analysis	Mainly Europe and N America	47 studies N=25,157	1.05 (1.03, 1.07)	Never HRT use:1.10 (95% CI, 1.07-1.13); Ever HRT use: 0.95 (95% CI, 0.92-0.99)
PROSTATE (C61)					
Present study	Cohort	UK	5.2 million (9)	0.98 (0.95, 1.00)*	BMI≤27kg/m^2: RR 1.10 (1.05, 1.15)* BMI>27kg/m^2: RR 0.88 (0.84, 0.92)*
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	27 studies	1.03 (1.00, 1.07)	-
Discacciata 2012[21]	Meta-analysis	US, Europe, Australia, Japan	25 studies N=1.1 million	N/A	Localised disease: 0.94 (0.91, 0.97); Advanced disease: 1.09 (1.02, 1.16)
KIDNEY (C64)					
Present study	Cohort	UK	5.2 million (9)	1.25 (1.17, 1.33)*	No evidence of non-linearity or effect modification
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	17 studies	Not reported	Men: 1.24 (0.95, 1.62); Women: 1.34 (1.25, 1.43)
Song 2008[2]	Cohort	Korea	170,481 (9)	N/A	Post-menopausal women only: 1.46 (1.10, 2.01)**
Ildaphonse 2009[22]	Meta-analysis	Mostly Europe and US	28 studies		Women only: 1.34 (1.28, 1.40)**
BLADDER (C67)					
Present study	Cohort	UK	5.2 million (9)	1.03 (0.99, 1.06)*	No evidence of non-linearity or effect modification
Reeves 2007[15]	Cohort	UK	1.2 million	-	Women aged >50 only: 1.04 (0.94, 1.16)**
BRAIN/CNS (C71-72)					
Present study	Cohort	UK	5.2 million (9)	1.04 (0.99, 1.10)*	No evidence of non-linearity or effect modification
Benson 2008[23]	Cohort	UK	1.3 million (6)	N/A	Women aged 50-64 at baseline only: 1.08 (1.01, 1.16)**

THYROID (C73)

Present study	Cohort	UK	5.2 million (9)	1.09 (1.00, 1.19)*	No evidence of non-linearity or effect modification
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	5 studies	Not reported	Men: 1.33 (1.04, 1.70); Women: 1.14 (1.06, 1.23)
Song 2008[2]	Cohort	Korea	170,481 (9)	N/A	Post-menopausal women only: 1.10 (0.90, 1.28)**
Clavel-Chapelon 2010[24]	Cohort	France	91,909 (13)	N/A	Women only: 1.20 (1.04, 1.38); some evidence of non-linearity with effect restricted to women with BMI over the median (22kg/m^2); no evidence of effect modification by smoking
Kitahara 2011[25]	Cohort	US	848,932 (10)	Not reported	Men: 1.21 (0.97, 1.49); Women: 1.16, (1.08, 1.24); some evidence of larger association among never smokers compared with current/former; no effect modification by age, education, alcohol, physical activity.

NON-HODGKIN LYMPHOMA (C82-85)

Present study	Cohort	UK	5.2 million (9)	1.03 (0.99, 1.06)*	No evidence of non-linearity or effect modification
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	9 studies	Not reported	Men: 1.06 (1.03, 1.09); Women: 1.07 (1.00, 1.14)
Song 2008[2]	Cohort	Korea	170,481 (9)	N/A	Post-menopausal women only: 1.05 (0.82, 1.40)**
Pylypchuk 2009[26]	Cohort	Netherlands	120,852 (13)	N/A	Diffuse large B-cell lymphoma: 1.26 (0.98, 1.62); Follicular lymphoma: 0.97 (0.67, 1.41); RR for combined NHL not reported
Kanda 2010[27]	Cohort	Japan	94,547 (14)	1.10 (0.86, 1.40)	-

MULTIPLE MYELOMA (C90)

Present study	Cohort	UK	5.2 million (9)	1.03 (0.98., 1.09)*	No evidence of non-linearity or effect modification
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	10 studies	Not reported	Men: 1.11 (1.05, 1.18); Women: 1.11 (1.07, 1.15)
Wallin 2011[28]	Meta-analysis	Mainly Europe and N America	20 studies, N=7.1 million	1.12 (1.08, 1.13)	No evidence of effect modification by gender, region, measured vs self-reported BMI, length of follow-up
Hofmann 2013[29]	Cohort	US	485,049	1.10 (1.00, 1.22)	No evidence of effect modification by gender, physical activity, age

LEUKAEMIA (C91-95)

Present study	Cohort	UK	5.2 million (9)	1.09 (1.05, 1.13)*	No evidence of non-linearity or effect modification
Renehan 2008[1]	Meta-analysis	Europe/N America/Asia-Pacific	7 studies	Not reported	Men: 1.08 (1.02, 1.14); Women: 1.17 (1.04, 1.32)
Song 2008[2]	Korea	Cohort	170,481 (9)	N/A	Post-menopausal women only: 1.54 (1.10, 2.10)**

Note: for cancer of the oral cavity, no relevant literature were found that met the systematic review inclusion criteria. Confidence intervals are 95% except where starred: * = (99% CI); **Converted to "per 5kg/m^2" from presented results (e.g. from per 1kg/m^2 or per standard deviation); NHL = Non-Hodgkin Lymphoma

Reference list for Table W3.1

1. 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**(9612): 569-78.
2. Song YM, Sung J, Ha M. Obesity and risk of cancer in postmenopausal Korean women. *J Clin Oncol* 2008; **26**(20): 3395-402.
3. Harriss DJ, Atkinson G, George K, et al. Lifestyle factors and colorectal cancer risk (1): systematic review and meta-analysis of associations with body mass index. *Colorectal Dis* 2009; **11**(6): 547-63.
4. Laake I, Thune I, Selmer R, Tretli S, Slattery ML, Veierod MB. A prospective study of body mass index, weight change, and risk of cancer in the proximal and distal colon. *Cancer Epidemiol Biomarkers Prev* 2010; **19**(6): 1511-22.
5. Bassett JK, Severi G, English DR, et al. Body size, weight change, and risk of colon cancer. *Cancer Epidemiol Biomarkers Prev* 2010; **19**(11): 2978-86.
6. 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**(10): 1127-39.
7. Renehan AG, Flood A, Adams KF, et al. Body mass index at different adult ages, weight change, and colorectal cancer risk in the National Institutes of Health-AARP Cohort. *Am J Epidemiol* 2012; **176**(12): 1130-40.
8. Semmens EO, Kopecky KJ, Grant E, et al. Relationship between anthropometric factors, radiation exposure, and colon cancer incidence in the Life Span Study cohort of atomic bomb survivors. *Cancer Causes Control* 2013; **24**(1): 27-37.
9. Tanaka K, Tsuji I, Tamakoshi A, et al. Obesity and liver cancer risk: an evaluation based on a systematic review of epidemiologic evidence among the Japanese population. *Jpn J Clin Oncol* 2012; **42**(3): 212-21.
10. Aune D, Greenwood DC, Chan DS, et al. Body mass index, abdominal fatness and pancreatic cancer risk: a systematic review and non-linear dose-response meta-analysis of prospective studies. *Ann Oncol* 2012; **23**(4): 843-52.
11. Suzuki R, Orsini N, Saji S, Key TJ, Wolk A. Body weight and incidence of breast cancer defined by estrogen and progesterone receptor status--a meta-analysis. *Int J Cancer* 2009; **124**(3): 698-712.
12. Bjorge T, Lukanova A, Jonsson H, et al. Metabolic syndrome and breast cancer in the me-can (metabolic syndrome and cancer) project. *Cancer Epidemiol Biomarkers Prev* 2010; **19**(7): 1737-45.
13. 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**(3): R76.
14. 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**(2): 330-5.
15. 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**(7630): 1134.
16. Lindemann K, Vatten LJ, Ellstrom-Engh M, Eskild A. Body mass, diabetes and smoking, and endometrial cancer risk: a follow-up study. *Br J Cancer* 2008; **98**(9): 1582-5.
17. 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**(12): 3119-30.
18. Yang TY, Cairns BJ, Allen N, et al. Postmenopausal endometrial cancer risk and body size in early life and middle age: prospective cohort study. *Br J Cancer* 2012; **107**(1): 169-75.
19. Chionh F, Baglietto L, Krishnan K, et al. Physical activity, body size and composition, and risk of ovarian cancer. *Cancer Causes Control* 2010; **21**(12): 2183-94.
20. Collaborative Group on Epidemiological Studies of Ovarian C. Ovarian cancer and body size: individual participant meta-analysis including 25,157 women with ovarian cancer from 47 epidemiological studies. *PLoS Medicine / Public Library of Science* 2012; **9**(4): e1001200.
21. 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**(7): 1665-71.
22. Ildaphonse G, George PS, Mathew A. Obesity and kidney cancer risk in men: a meta-analysis (1992-2008). *Asian Pacific Journal of Cancer Prevention: Apjcp* 2009; **10**(2): 279-86.
23. Benson VS, Pirie K, Green J, Casabonne D, Beral V. Lifestyle factors and primary glioma and meningioma tumours in the Million Women Study cohort. *Br J Cancer* 2008; **99**(1): 185-90.
24. Clavel-Chapelon F, Guillas G, Tondeur L, Kernaleguen C, Boutron-Ruault MC. Risk of differentiated thyroid cancer in relation to adult weight, height and body shape over life: the French E3N cohort. *Int J Cancer* 2010; **126**(12): 2984-90.
25. Kitahara CM, Platz EA, Freeman LE, et al. Obesity and thyroid cancer risk among U.S. men and women: a pooled analysis of five prospective studies. *Cancer Epidemiol Biomarkers Prev* 2011; **20**(3): 464-72.
26. 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**(3): 297-307.
27. Kanda J, Matsuo K, Inoue M, et al. Association of anthropometric characteristics with the risk of malignant lymphoma and plasma cell myeloma in a Japanese population: a population-based cohort study. *Cancer Epidemiol Biomarkers Prev* 2010; **19**(6): 1623-31.
28. Wallin A, Larsson SC. Body mass index and risk of multiple myeloma: a meta-analysis of prospective studies. *Eur J Cancer* 2011; **47**(11): 1606-15.
29. Hofmann JN, Moore SC, Lim U, et al. Body mass index and physical activity at different ages and risk of multiple myeloma in the NIH-AARP diet and health study. *Am J Epidemiol* 2013; **177**(8): 776-86.