Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies

*Endogenous Hormones and Breast Cancer Collaborative Group**

Summary

Background Associations between circulating concentrations of oestrogens, progesterone, and androgens with breast cancer and related risk factors in premenopausal women are not well understood. We aimed to characterise these associations with a pooled analysis of data from seven studies.

Methods Individual participant data for prediagnostic sex hormone and sex hormone-binding globulin (SHBG) concentrations were contributed from seven prospective studies. We restricted analyses to women who were premenopausal and younger than 50 years at blood collection, and to women with breast cancer diagnosed before age 50 years. We estimated odds ratios (ORs) with 95% CIs for breast cancer associated with hormone concentrations by conditional logistic regression in cases and controls matched for age, date of blood collection, and day of cycle, with stratification by study and further adjustment for cycle phase. We examined associations of hormones with risk factors for breast cancer in control women by comparing geometric mean hormone concentrations in categories of these risk factors, adjusted for study, age, phase of menstrual cycle, and body-mass index (BMI). All statistical tests were two-sided.

Findings We included data for up to 767 women with breast cancer and 1699 controls in the risk analyses. Breast cancer risk was associated with a doubling in concentrations of oestradiol (OR 1·19, 95% CI 1·06–1·35), calculated free oestradiol (1·17, 1·03–1·33), oestrone (1·27, 1·05–1·54), androstenedione (1·30, 1·10–1·55), dehydroepiandrosterone sulphate (1·17, 1·04–1·32), testosterone (1·18, 1·03–1·35), and calculated free testosterone (1·08, 0·97–1·21). Breast cancer risk was not associated with luteal phase progesterone (doubling in concentration OR 1·00, 95% CI 0·92–1·09), and adjustment for other factors had little effect on any of these ORs. Cross-sectional analyses in control women showed several associations of sex hormones with breast cancer risk factors.

Interpretation Circulating oestrogens and androgens are positively associated with the risk for breast cancer in premenopausal women.

Funding Cancer Research UK.

Lancet Oncol 2013; 14: 1009–19
Published Online July 24, 2013
http://dx.doi.org/10.1016/S1470-2045(13)70301-2

*Members listed at end of paper

Correspondence to:
Prof Timothy Key, Endogenous Hormones and Breast Cancer Collaborative Group, Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Richard Doll Building, Roosevelt Drive, Oxford OX3 7LF, UK
tim.key@ceu.ox.ac.uk

Introduction

Risk of breast cancer is affected by several reproductive and hormonal factors and endogenous sex hormones are also thought to influence risk.[1] Sufficient data now exist from studies of hormones and breast cancer in postmenopausal women to show that risk is positively associated with circulating concentrations of oestrogens and androgens,[2–4] but fewer data are available for premenopausal women and hormone measurements are complicated by the substantial variation in hormone concentrations across the menstrual cycle.

The Endogenous Hormones and Breast Cancer Collaborative Group was established to undertake pooled analyses of individual data from prospective studies to increase the precision of the estimated associations of endogenous hormones with the risk of breast cancer.[2] We report here a collaborative analysis of data from seven studies. We describe the associations of circulating concentrations of sex hormones with breast cancer risk, including examination of consistency between studies, associations in subgroups, and the effects of adjustment for other risk factors. We also describe cross-sectional analyses of the associations of circulating sex hormones and sex hormone-binding globulin (SHBG) with risk factors for breast cancer. Our aim was to improve understanding of the role of premenopausal sex hormones in breast cancer diagnosed before menopause, because hormonal changes after menopause will probably influence the association of premenopausal hormone levels with the risk for postmenopausal breast cancer. We therefore restricted all analyses to women who were premenopausal and younger than 50 years at blood collection, and to case–control sets in which the case patient was diagnosed with breast cancer before age 50 years.

Methods

Data collection

Published studies were eligible for the collaborative reanalysis if they included data for endogenous hormones and breast cancer risk from prospectively collected blood samples of premenopausal women.

We included the following studies, which were identified by computer-aided literature searches and through discussions with colleagues: CLUE I, Washington County, MD, USA;[5,6] Columbia, MO, USA;[7] the European Prospective Investigation into Cancer and Nutrition (EPIC), Europe;[8] Guernsey, UK;[9] Nurses' Health Study II (NHS-II), USA;[10–12] New York University Women's Health Study (NYU WHS), USA;[13] and the Study of Hormones and Diet in the Etiology of Breast Tumors (ORDET), Italy.[14] Most women in these studies were of white European ethnic origin. Two additional studies in the Collaborative Group had prospective hormone data but were not included in the analyses reported here because data for day of menstrual cycle at blood collection were not available: the Melbourne Collaborative Cohort Study, Australia,[15] and Radiation Effects Research Foundation study phases 1 and 2, Japan.[16,17] One final study[18] was identified but the data could not be retrieved; this study included 17 patients with breast cancer and 67 matched controls in women who were premenopausal at blood collection.

See Online for appendix

Table 1 summarises the designs of included studies. Participants in NHS-II were volunteers who were nurses, and in the other cohorts the participants were volunteers living in areas near the recruitment centres. Details of the recruitment of participants, informed consent, ethics approvals, and definitions of reproductive variables are reported in the original publications. The present analyses did not need further ethics approval. Most cases were of invasive breast cancer, but five of the studies (Columbia, EPIC, NHS-II, NYU WHS, and ORDET) also included some in-situ cases. Cases were individually matched to between two and four controls: all studies matched participants on age and date of blood sample (or follow-up time for EPIC), and on the day or phase of menstrual cycle at blood collection. Collaborators were asked to provide data for concentrations of the hormones oestradiol (total), oestrone, progesterone, androstenedione, dehydroepiandrosterone sulphate (DHEAS), testosterone, and SHBG, and data for reproductive, anthropometric, and lifestyle factors for each woman in their study, when available. Women who used hormonal contraceptives or other exogenous sex hormones at the time of blood collection were excluded, as were women missing information for date of birth, date at blood collection, day of menstrual cycle at blood collection, or date of diagnosis (cases).

The appendix contains brief details of the assays used, with further details in the original publications. Six studies measured hormone concentrations in serum whereas one (NHS-II) used heparin plasma; for convenience we refer to serum concentrations for the

	Recruitment population	Recruitment period	Fasting status	Storage temperature	Matching criteria				
					Controls per case	Age at blood collection	Date of blood sample	Day of cycle	Other matching criteria
CLUE I, USA[5,6]	Residents of Washington County, MD, USA	1974	Non-fasting	–70°C	2	±1 year	±14 days	±1 day	Time of day, fasting status, ethnic group, freeze–thaw history of serum sample
Columbia, USA[7]	Residents of Columbia, MO, USA	1977–89	Non-fasting	–70°C	2	±2 years	±1 year	±2 days	Time of day at blood collection
EPIC, Europe[8]	Volunteers in Denmark, France, Germany, Greece, Italy, Netherlands, Spain, Sweden, and UK	1992–98	Matched	–196°C*	2	±6 months	No (incidence density sampling)	Five phases	Time of day at blood collection, fasting status, subcohort
Guernsey, UK[9]	Residents of the island of Guernsey, UK	1977–90	Non-fasting	–20°C	3	±2 years	±1 year	±1 day	..
Nurses' Health Study II phases 1 (1999–2003 follow-up cycles) and 2 (2005–09 follow-up cycles), USA[10–12]	Registered nurses in the USA	1996–99	Matched	–130°C	2	±2 years	±2 months	±1 day for luteal blood sample†	Time of day, fasting status
NYU WHS, USA[13]	Women attending a breast cancer screening centre in New York, NY, USA	1985–91	Non-fasting	–80°C	2	±6 months	±3 months	5 phases and day	Number of subsequent samples
ORDET, Italy[14]	Residents of Varese province, Italy	1987–92	12 h fast before collection; samples taken 0730–0900 h	–80°C	4	±5 years	±89 days	All days 20–24	Daylight saving period, recruitment centre

CLUE I=Washington County, MD Study "Give us a clue to cancer and heart disease". EPIC=European Prospective Investigation into Cancer and Nutrition. NYU WHS=New York University Women's Health Study. ORDET=Study of Hormones and Diet in the Etiology of Breast Tumors. *Most samples were stored in liquid nitrogen at –196°C, apart from in Denmark where they were stored in nitrogen vapour at –150°C. †Patients were asked to provide follicular sample at 3–5 days and luteal sample at 7–9 days before anticipated start of next cycle.

Table 1: **Prospective cohort studies combined in the collaborative reanalysis**

pooled analyses. We calculated circulating concentrations of free oestradiol and free testosterone from the concentrations of oestradiol and testosterone and of SHBG according to the law of mass action, with albumin assumed to be constant (40 g/L).[19,20]

Statistical analysis

We grouped day of cycle at blood collection into six categories, according to the number of days until next period if available (backward dating), otherwise according to days since last period (forward dating). The six categories were as follows: early follicular (day ≥24 backwards or days 1–5 forwards), late follicular (19–23 backwards or 6–10 forwards), mid-cycle (15–18 backwards or 11–14 forwards), early luteal (11–14 backwards or 15–18 forwards), mid luteal (5–10 backwards or 19–24 forwards), or late luteal (0–4 backwards or ≥25 forwards). For CLUE I and Columbia, day of cycle was assessed with forward dating

	Women*	Oestradiol, pmol/L	Calculated free oestradiol, pmol/L	Oestrone, pmol/L	Luteal phase progesterone, nmol/L	Androstenedione, nmol/L	DHEAS, nmol/L	Testosterone, nmol/L	Calculated free testosterone, pmol/L	SHBG, nmol/L
CLUE I, USA[5,6]										
Cases	21	172 (134–222)	2·03 (1·55–2·64)	252 (211–301)	5·32 (1·73–16·3)	2·98 (2·31–3·85)	3903 (2787–5465)	..	..	69·9 (58·0–84·2)
Controls	42	168 (137–206)	1·85 (1·51–2·26)	239 (207–275)	9·62 (5·47–16·9)	2·88 (2·42–3·43)	3853 (3023–4910)	..	..	74·3 (65·3–84·4)
Columbia, USA[7]										
Cases	13	239 (165–347)	3·26 (2·24–4·75)	..	..	..	..	1·00 (0·79–1·28)	13·7 (9·86–19·1)	48·2 (34·3–67·7)
Controls	24	316 (257–387)	4·05 (3·34–4·92)	..	..	..	..	0·86 (0·73–1·02)	10·7 (9·10–12·7)	56·6 (48·3–66·4)
EPIC, Europe[8]										
Cases	206	318 (285–355)	4·60 (4·13–5·12)	384 (354–416)	8·42 (6·30–11·3)	5·59 (5·22–5·98)	3712 (3469–3972)	1·70 (1·60–1·81)	25·2 (23·2–27·3)	43·5 (40·6–46·6)
Controls	408	296 (275–318)	4·25 (3·94–4·60)	360 (339–383)	12·3 (9·84–15·4)	4·92 (4·68–5·18)	3341 (3169–3522)	1·56 (1·49–1·63)	23·3 (21·8–24·8)	43·0 (40·9–45·3)
Guernsey, UK[9]										
Cases	32	323 (253–412)	3·16 (2·39–4·17)	..	10·7 (5·84–19·5)	..	2253 (1410–3599)	1·17 (0·97–1·40)	13·2 (11·3–15·5)	68·6 (59·5–79·1)
Controls	94	282 (246–323)	3·02 (2·52–3·62)	..	10·6 (7·25–15·4)	..	2548 (1924–3375)	1·12 (1·02–1·23)	13·4 (11·6–15·5)	61·5 (55·8–67·7)
Nurses' Health Study II phase 1, USA[10,11]										
Cases	139	182 (166–199)	2·30 (2·12–2·49)	150 (142–159)	45·7 (41·1–50·8)	3·91 (3·68–4·16)	2302 (2129–2489)	0·92 (0·87–0·99)	11·3 (10·5–12·2)	57·9 (53·8–62·3)
Controls	268	164 (153–177)	2·08 (1·95–2·22)	145 (138–151)	43·5 (39·9–47·4)	3·89 (3·72–4·06)	2208 (2089–2333)	0·90 (0·86–0·94)	10·9 (10·3–11·5)	58·5 (55·5–61·8)
Nurses' Health Study II phase 2, USA[12]										
Cases	105	193 (175–213)	2·21 (2·02–2·42)	161 (150–173)	40·7 (34·6–47·9)	..	2838 (2556–3151)	0·91 (0·85–0·98)	9·6 (8·7–10·5)	70·6 (65·3–76·3)
Controls	203	186 (174–199)	2·25 (2·11–2·40)	163 (154–171)	38·1 (33·9–42·9)	..	2642 (2449–2851)	0·91 (0·87–0·96)	10·6 (10·0–11·3)	62·4 (59·0–66·0)
NYU WHS phase 2, USA[13]										
Cases	137	..	..	..	..	4·30 (3·96–4·67)	3978 (3625–4366)	1·01 (0·91–1·12)	14·0 (12·4–15·8)	48·1 (44·1–52·4)
Controls	258	..	..	..	..	4·07 (3·83–4·33)	3869 (3598–4161)	0·95 (0·88–1·03)	13·1 (11·9–14·3)	47·8 (44·8–51·0)
ORDET, Italy[14]										
Cases	84	300 (274–329)	3·66 (3·34–4·00)	..	38·2 (32·7–44·6)	5·26 (4·38–6·32)	3856 (3153–4715)	0·85 (0·75–0·97)	9·9 (8·5–11·6)	62·0 (56·6–68·0)
Controls	336	282 (259–306)	3·50 (3·23–3·79)	..	32·4 (28·3–37·1)	5·79 (5·38–6·23)	3921 (3604–4265)	0·84 (0·79–0·90)	10·1 (9·3–10·8)	59·8 (57·1–62·6)

Geometric mean hormone concentrations for Nurses' Health Study II are obtained using the follicular phase data for oestradiol, calculated free oestradiol and oestrone and using the luteal phase data for all other hormones. DHEAS=dehydroepiandrosterone sulphate. SHBG=sex hormone-binding globulin. CLUE I=Washington County, MD Study "Give us a clue to cancer and heart disease". EPIC=European Prospective Investigation into Cancer and Nutrition. NYU WHS=New York University Women's Health Study. ORDET=Study of Hormones and Diet in the Etiology of Breast Tumors. ..=data not available. *Women with known phase of cycle and values for oestradiol (apart from NYU WHS, in which numbers are for women with values for testosterone).

Table 2: **Geometric mean hormone concentrations (with 95% CIs)**

for all participants. For Guernsey and NHS-II, day of cycle was determined backwards for all participants (apart from one case in NHS-II). For the other three studies, the percentages of patients determined with backward dating (otherwise forward dating) were as follows: 54·4% for cases and 51·0% for controls in EPIC; 75·2% for cases and 82·9% for controls in NYU WHS; and 94·0% for cases and 96·1% for controls in ORDET.

In NHS-II, participants provided two blood samples at baseline, one collected in the follicular phase and one in the luteal phase. For most of the analyses reported here we use values for oestradiol and oestrone from the follicular phase, and progesterone, androstenedione, DHEAS, testosterone, and SHBG from the luteal phase; in the analyses of oestradiol and free oestradiol subdivided by phase of cycle we used both the follicular and the luteal measures, as appropriate. In the other studies, only one blood sample was collected from each participant.

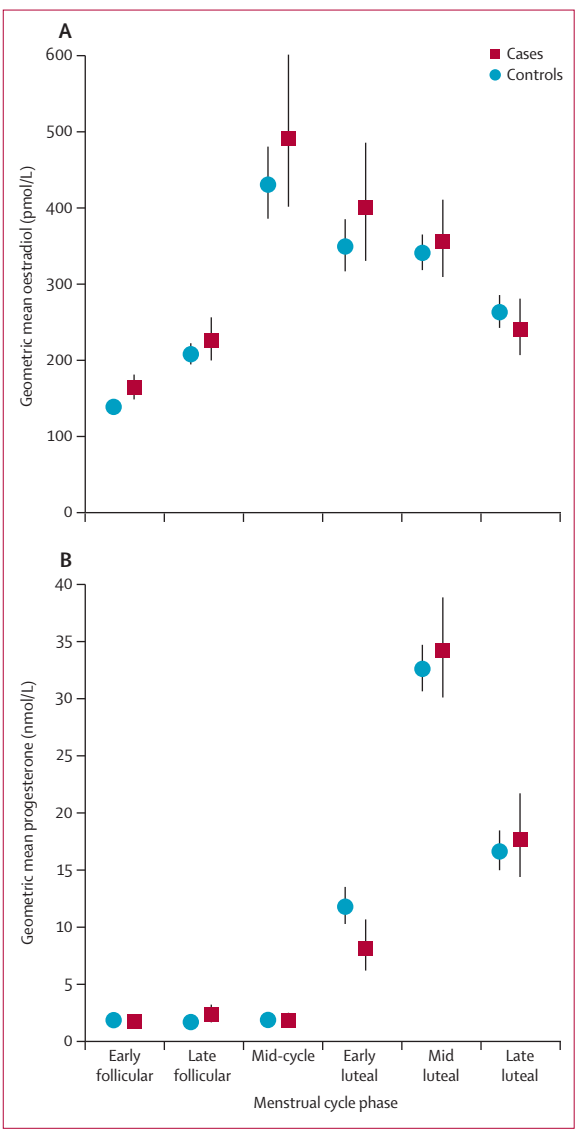

Figure 1: **Geometric mean oestradiol concentrations (A) and progesterone concentrations (B), by phase of menstrual cycle**
Adjusted for study and age at blood collection. Vertical lines show 95% CIs.

All women were classified as premenopausal in the contributed datasets, with the criteria for this based on questionnaire information, as described in the original studies; three studies additionally measured serum follicle-stimulating hormone (FSH) concentration and excluded women with FSH values higher than the cutoff recommended by their laboratory (Guernsey, NYU WHS, and ORDET). We restricted analyses to cases diagnosed before age 50 years (and their matched controls), so that most cases would have been diagnosed when premenopausal; this restriction further served to reduce the possibility that some participants were peri-menopausal at blood collection.

All included studies used a nested case–control design, with assays arranged so that case–control sets were generally measured in the same batch, thus eliminating interassay variation from the case–control comparisons. We retained the original matched sets in the analyses. Some studies used density sampling, meaning that an individual participant could appear more than once in a data file.

We used conditional logistic regression to estimate the odds ratio (OR) for breast cancer in relation to the serum concentrations of hormones and SHBG, categorising women in each study according to the quintiles of hormone concentration for the controls in that study, after standardisation for phase of menstrual cycle with residuals from the study-specific mean for each cycle phase; for progesterone, we restricted the analysis to samples collected in the luteal phase. For each study, we fitted the simple linear regression model: $\log(\text{hormone}) = A + B \times (\text{phase of cycle})$, where phase of cycle was a categorical variable. The residuals from this model were then added to the mean log(hormone) value and the result exponentiated to give the standardised value. We used study-specific quintile cut-points because the absolute concentrations of hormones and SHBG vary between studies due to laboratory variation; further explanation of this approach is provided in previous publications.[2,21] To test for the significance of the association and to provide a summary measure of risk we also estimated the OR associated with a unit increase in a continuous variable equal to the logarithm to the base 2 of the hormone concentration. A unit increase in this variable is equivalent to a doubling in hormone concentration. Heterogeneity in linear trends between studies was tested by comparing the χ^2 values for models with and without a (study) × (linear trend) interaction term. We also used χ^2 tests to examine whether evidence of heterogeneity existed in the associations of hormones with risk of breast cancer according to subgroups defined according to years from blood collection to diagnosis (<4 years or ≥4 years), stage

of disease (in situ or invasive), oestrogen receptor status (positive or negative), progesterone receptor status (positive or negative), HER2 receptor status (positive or negative), phase of menstrual cycle at blood collection (apart from for progesterone; follicular, mid-cycle, or luteal), age at menarche (<14 years or ≥14 years), parity (nulliparous or parous), age at first full-term pregnancy (<25 years or ≥25 years), mother or sister with breast cancer (no or yes), body-mass index (BMI; <25 kg/m² or ≥25 kg/m²), smoking (never or past, current), alcohol intake at recruitment (<10 g per day or ≥10 g per day), previous use of hormonal contraceptives (no or yes), and assay method for oestradiol, calculated free oestradiol, oestrone, testosterone, and calculated free testosterone (extraction or non-extraction). We also investigated the associations of hormones with breast cancer risk after adjustment for reproductive and hormonal risk factors for breast cancer: age at menarche (<12 years, 12–13 years, ≥14 years, or unknown); parity (zero, one, two, three, or four or more full-term pregnancies, or unknown); age at first full-term pregnancy (<20 years, 20–24 years, 25–29 years, ≥30 years, or unknown); BMI (<22·5 kg/m², 22·5–24·9 kg/m², 25·0–27·4 kg/m², 27·5–29·9 kg/m², ≥30·0 kg/m², or unknown).

Concentrations of the hormones and SHBG were positively skewed, we therefore used log-transformed concentrations for all parametric analyses. We examined associations of hormones with risk factors for breast cancer in the controls; the numbers of controls in these analyses were somewhat larger than the numbers in the risk analyses because all controls who were pre-menopausal and aged younger than 50 years at blood collection were eligible, irrespective of the age at blood collection or diagnosis of their matched case. We calculated geometric means and 95% CIs according to categories of these factors, adjusting for study, age (<40 years, 40–44 years, and 45–49 years), cycle phase, and BMI, as appropriate. We used *F* tests to assess heterogeneity in the geometric mean hormone concentrations between the categories of risk factors, and where appropriate to test for trends across the categories, with the ordered categories scored from 1 to the maximum number of categories. The heterogeneity between studies in the associations of hormones with risk factors was assessed by adding a (study) × (factor) interaction term to the model and using the *F* test to calculate its significance.

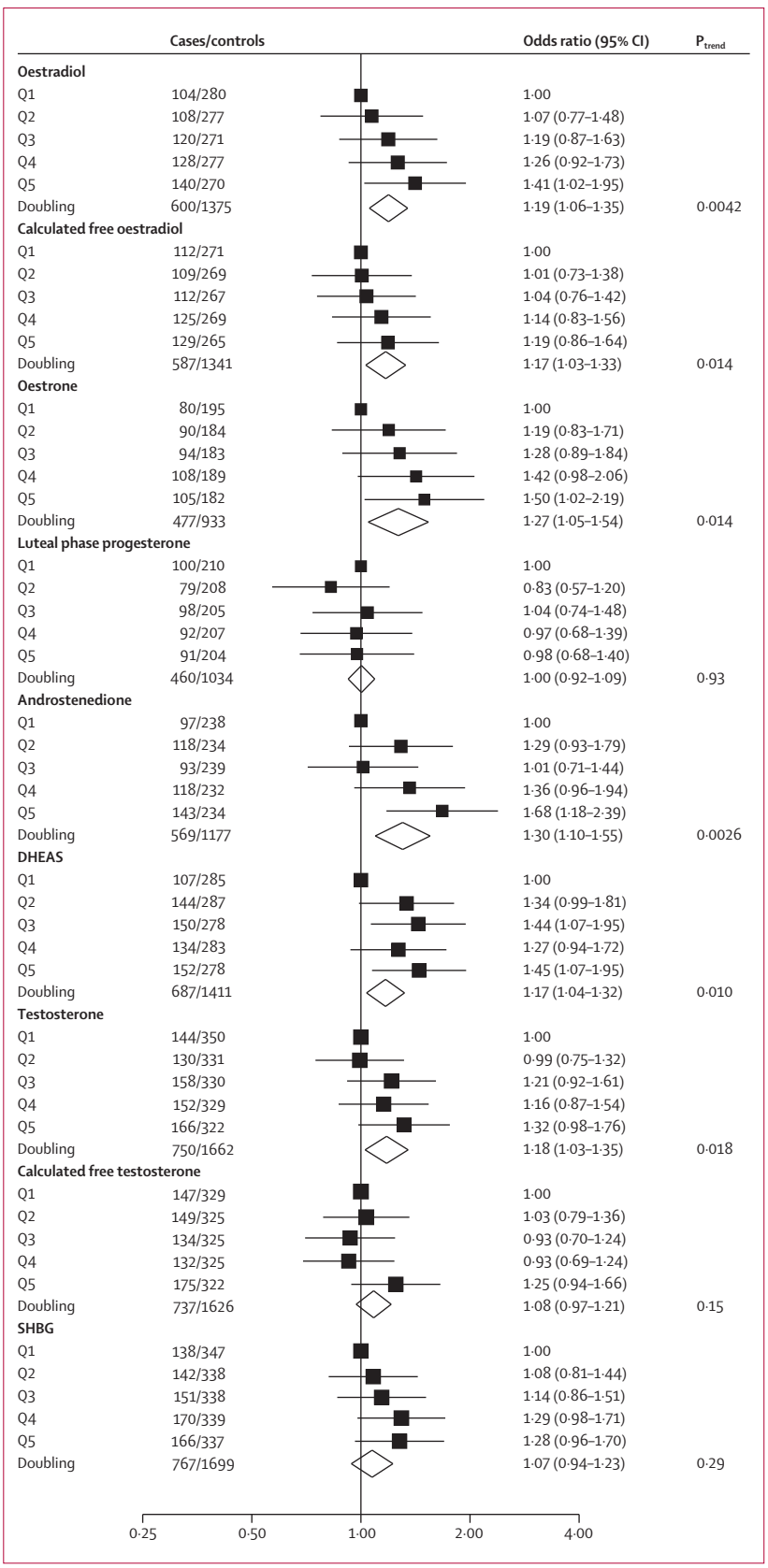

Figure 2: **Odds ratios for breast cancer associated with sex hormones and SHBG**
Black squares show odds ratios in quintiles (Q1–Q5; study-specific fifths after adjustment for phase of cycle within each study), and the horizontal lines show 95% CIs. The area of each square is proportional to the amount of statistical information (inverse of the variance of the logarithm of the odds ratios). The diamonds show the odds ratio for a doubling in concentration, and the widths of the diamonds show the 95% CIs. Estimates are from conditional logistic regression on case–control sets matched within each study. DHEAS=dehydroepiandrosterone sulphate. SHBG=sex hormone-binding globulin.

All statistical tests were two-sided and we set statistical significance at the 5% level. All analyses were done with Stata version 12.0.

Role of the funding source

The sponsor of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The members of the writing committee had full access to all data in the study and had the final responsibility for the decision to submit the manuscript for publication.

Results

We included 767 women with breast cancer and 1699 matched controls from the seven studies in the risk analyses; all these women had data on SHBG, with smaller numbers for the individual sex hormones. Mean age at blood collection ranged from 35·6 years for cases in CLUE to 42·2 years for cases and controls in EPIC (appendix p 4). The median time between blood collection and diagnosis ranged from 2 years in EPIC to 9 years in CLUE. Geometric mean concentrations of sex hormones and SHBG in controls ranged from 164 pmol/L to 316 pmol/L for oestradiol, 1·85 pmol/L to 4·25 pmol/L for calculated free oestradiol, 145 pmol/L to 360 pmol/L for oestrone, 9·62 nmol/L to 43·5 nmol/L for luteal phase progesterone, 2·88 nmol/L to 5·79 nmol/L for androstenedione, 2208 nmol/L to 3921 nmol/L for DHEAS, 0·84 nmol/L to 1·56 nmol/L for testosterone, 10·1 pmol/L to 23·3 pmol/L for calculated free testosterone, and 43·0 nmol/L to 74·3 nmol/L for SHBG (table 2).

Figure 1 shows geometric mean concentrations of oestradiol and progesterone in cases and controls by phase of menstrual cycle at blood collection. For oestradiol, geometric mean values were higher, although not statistically significantly so in cases than in controls at cycle phases apart from the late luteal phase. For progesterone, geometric means were lower in cases than controls in the early luteal phase, with small non-significant differences in the other phases.

Concentrations of oestradiol, calculated free oestradiol, oestrone, androstenedione, DHEAS, and testosterone were positively associated with breast cancer risk (figure 2). Concentrations of luteal phase progesterone, calculated free testosterone, and SHBG were not significantly associated with such risk. In a sensitivity analysis restricted to women with blood collected at ages younger than 45 years and breast cancer diagnosed before age 45 years the results showed similar trends (appendix p 5). We noted no significant heterogeneity between studies in the associations of these hormones with breast cancer risk (appendix pp 6–14). Further adjustment for age at menarche, age at first full-term pregnancy, number of full-term pregnancies, and BMI did not substantially change the ORs, apart from that after adjustment we noted a significant positive association of calculated free testosterone with risk (OR for a doubling was 1·14, 95% CI 1·01–1·28; p_{trend}=0·031; appendix p 15).

Appendix pp 16–24 show results of the subgroup analyses. Four of 130 tests for heterogeneity were significant in our subgroup analyses of the nine hormones: for oestradiol the OR for a doubling in concentration was 1·26 (95% CI 1·10–1·44) for never or past smokers and 0·94 (0·75–1·18) for current smokers ($p_{heterogeneity}$=0·034); 1·01 (0·82–1·24) for never users and 1·32 (1·14–1·53) for past users of hormonal contraceptives ($p_{heterogeneity}$=0·030); for oestrone the OR for a doubling in concentration was 1·74 (0·99–3·03) for progesterone receptor-positive cancers and 0·54 (0·27–1·08) for progesterone receptor-negative cancers ($p_{heterogeneity}$=0·010); and for luteal phase progesterone the OR for a doubling in concentration was 1·25 (1·01–1·55) for nulliparous women and 0·99 (0·90–1·09) for parous women ($p_{heterogeneity}$=0·034).

Three subgroup analyses were of particular a-priori interest. No evidence suggested that any of the ORs varied by the time between blood collection and diagnosis (appendix pp 16–24). For oestrogens and androgens, the ORs were larger for oestrogen-receptor positive tumours, the ORs were larger for oestrogen receptor-positive tumours than they were for oestrogen receptor-negative tumours, but none of these differences were significant (table 3). For oestradiol according to phase of menstrual cycle the ORs for a doubling in concentration were 1·25 (95% CI 1·06–1·48) for follicular samples, 1·20 (0·81–1·79) for mid-cycle samples, and 1·13 (0·92–1·37) for luteal samples ($p_{heterogeneity}$=0·732; appendix p 16).

Compared with women with a BMI of less than 22·5 kg/m², women with a BMI of 30 kg/m² or more had lower mean concentrations of oestradiol (by 17%), luteal phase progesterone (by 28%), and SHBG (by 46%); conversely, we noted a positive association with BMI for mean concentrations of calculated free oestradiol (by

	Oestrogen receptor positive		Oestrogen receptor negative		$p_{heterogeneity}$
	Cases/controls	Odds ratio (95% CI)	Cases/controls	Odds ratio (95% CI)	
Oestradiol	147/374	1·25 (0·95–1·65)	71/209	1·09 (0·76–1·57)	0·56
Calculated free oestradiol	147/374	1·22 (0·91–1·63)	71/209	1·03 (0·68–1·54)	0·50
Oestrone	107/205	1·26 (0·77–2·06)	37/72	0·90 (0·45–1·82)	0·45
Luteal phase progesterone	152/369	1·05 (0·88–1·24)	67/184	1·13 (0·88–1·47)	0·62
Androstenedione	124/237	1·45 (0·98–2·15)	54/106	1·11 (0·58–2·14)	0·50
DHEAS	170/327	1·24 (0·97–1·57)	67/130	0·91 (0·62–1·34)	0·19
Testosterone	211/495	1·13 (0·88–1·43)	99/265	1·03 (0·76–1·39)	0·66
Calculated free testosterone	211/495	1·08 (0·88–1·33)	99/264	1·01 (0·78–1·30)	0·66
SHBG	214/503	1·04 (0·80–1·35)	102/271	1·08 (0·77–1·52)	0·86

Estimates are from conditional logistic regression on case–control sets matched within each study and adjusted for phase of menstrual cycle at blood collection within study. DHEAS=dehydroepiandrosterone sulphate. SHBG=sex hormone-binding globulin.

Table 3: **Odds ratios for breast cancer associated with a doubling in concentrations of hormones and SHBG, subdivided by oestrogen receptor status**

10%), oestrone (by 16%), DHEAS (by 8%), testosterone (by 7%), and calculated free testosterone (by 63%), with means adjusted for age, study, and cycle phase (figure 3).

Appendix pp 25–31 show the associations of sex hormones with age at blood collection, age at menarche, parity, family history of breast cancer, smoking, alcohol intake at recruitment, and previous use of hormonal contraceptives. Concentrations of sex hormones were lower in older women than in younger women, whereas SHBG was higher in older women (appendix p 25). Parity was inversely associated with calculated free testosterone, but was not associated with concentrations of the other sex hormones or SHBG (appendix p 27), and none of the hormones or SHBG was associated with age at menarche or family history of breast cancer (appendix pp 26, 28). Compared with never-smokers, current smokers of at least 15 cigarettes per day had increased concentrations of androstenedione (by 21%), DHEAS (by 12%), testosterone (by 12%), and calculated free testosterone (by 13%; appendix p 29). Compared with women who did not consume alcohol, women with an alcohol intake of at least 20 g per day had increased concentrations of androstenedione (by 13%), DHEAS (by 16%), testosterone (by 23%), and calculated free testosterone (by 23%; appendix p 30). Further adjustment of the analyses by smoking for alcohol, and of the analyses by alcohol for smoking, had no material effect on the results (data not shown). Women who had previously used hormonal contraceptives had lower concentrations of oestradiol (by 7%), oestrone (by 7%), androstenedione (by 5%), and SHBG (by 4%) than did women who had not used hormonal contraceptives (appendix p 31).

Discussion

Our worldwide collaboration brought together and reanalysed individual participant data from seven studies of risk of breast cancer and endogenous sex hormones measured in prospectively collected blood samples (panel). All women were premenopausal and provided information about the phase of the menstrual cycle at the time of blood collection. Oestradiol, calculated free oestradiol, oestrone, androstenedione, DHEAS, and testosterone concentrations were positively associated with risk of breast cancer, whereas luteal phase progesterone and SHBG were not associated with risk. Calculated free testosterone was positively associated with breast cancer risk only after adjustment for reproductive factors and BMI. These associations did not vary according to the time between blood collection and diagnosis, making reverse causality unlikely, and (with the exception of calculated free

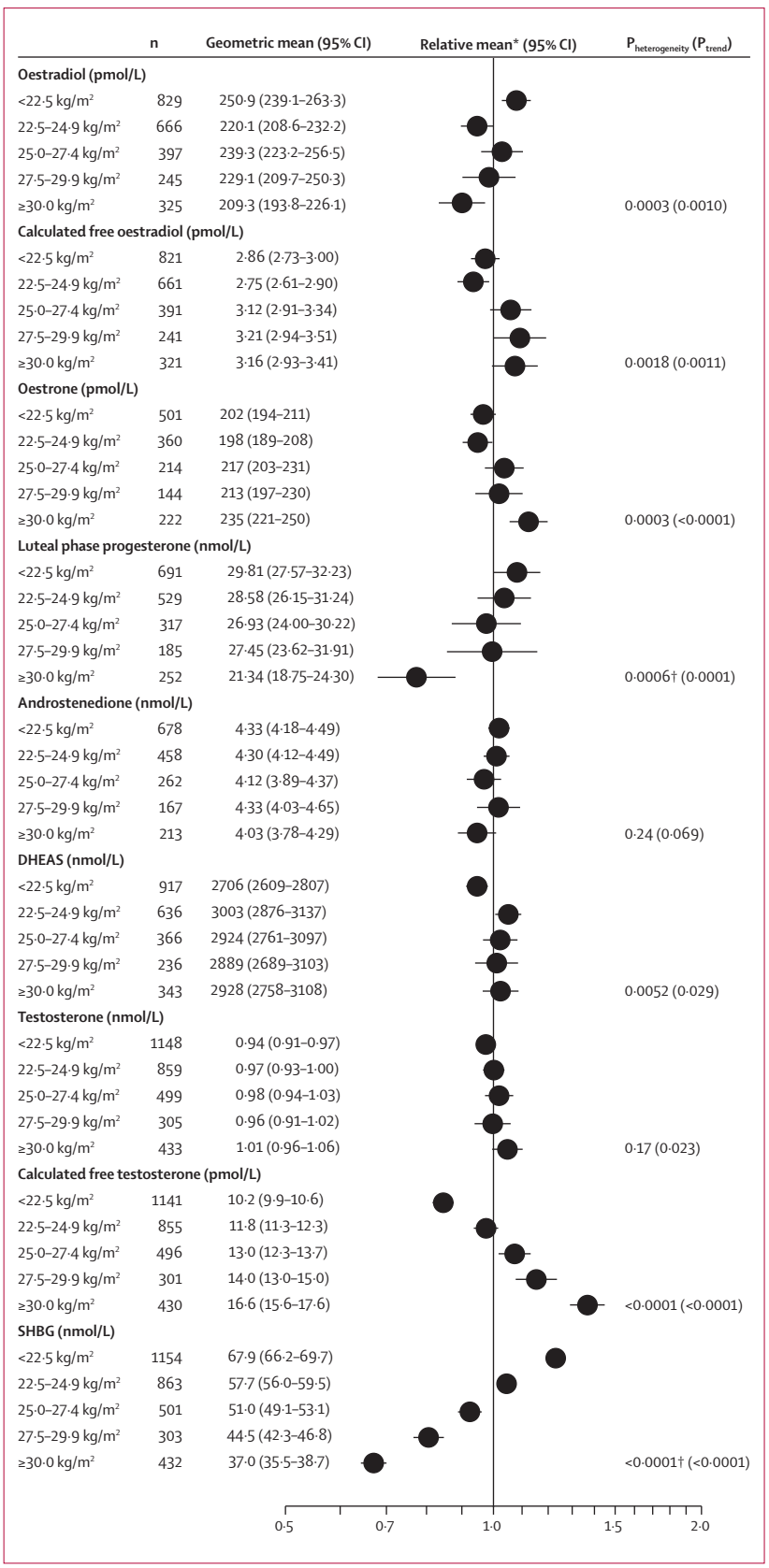

***Figure 3*: Geometric mean hormone and SHBG concentrations (with 95% CIs) in controls, by body-mass index**

Adjusted for study, age at blood collection, and phase of menstrual cycle. *Means scaled to overall mean concentration (vertical line). †Significant (p<0·05) interaction with study. DHEAS=dehydroepiandrosterone sulphate. SHBG=sex hormone-binding globulin.

Panel: **Research in context**

Systematic review

Literature on epidemiological studies of breast cancer was identified from electronic searches of PubMed between Jan 1, 1980, and June 30, 2012, with combinations of the search terms "breast cancer" and "endogenous hormones", supplemented by searching review articles, and discussions with colleagues. Eligible studies needed to have sex hormone concentrations measured in serum or plasma collected prospectively from women who subsequently developed breast cancer and from control women who did not develop breast cancer. In addition, studies needed to have information on the stage of the menstrual cycle at blood collection.

Interpretation

We included seven eligible studies and principal investigators contributed information from 767 women with breast cancer. We report on the relation of circulating concentrations of eight sex hormones and sex hormone-binding globulin (SHBG) with breast cancer risk, overall and by subgroups, including stage of disease, receptor status of tumours, and other risk factors for breast cancer. We also describe the associations of sex hormones with risk factors for breast cancer in the control women. Breast cancer risk was positively associated with circulating concentrations of oestrogens and androgens.

testosterone) were not materially affected by adjustment for other risk factors, suggesting that confounding is unlikely. These results therefore strongly suggest that breast cancer risk in premenopausal women increases with increasing concentrations of these sex hormones. The results are qualitatively similar to those reported in postmenopausal women, but smaller in magnitude.[2–4]

The analyses reported in this paper were all based on a one-off hormone measurement for each woman. Measurements of hormone concentrations are subject to random error associated with assay variation, and fluctuations in serum levels within individual women. Studies of the reproducibility of sex hormones in premenopausal women for up to 3 years have shown intraclass correlations of about 0·6 or above for androgens and SHBG, but correlations of about 0·4 or less for oestrogens and progesterone.[22,23] Therefore our reported associations between hormone concentrations and breast cancer risk are probably underestimates of the true associations, particularly for oestrogens, but more reproducibility data are required.

Our subgroup analyses showed heterogeneity in the associations of oestradiol with risk according to smoking and previous use of hormonal contraceptives, of oestrone with risk according to progesterone receptor status, and of luteal phase progesterone with risk according to parity, but there was no significant heterogeneity according to any other combination of risk factors and hormones. All the sex hormones had larger associations with the risk of oestrogen receptor-positive breast cancer than they did with the risk of oestrogen receptor-negative disease; these differences were not statistically significant, but study power was low because of the small numbers of cases with oestrogen receptor-negative disease (eg, 71 cases for oestradiol). Because we did 130 subgroup analyses, some of the four analyses that were nominally significant may have occurred from chance.

For oestradiol, the plot of geometric mean concentrations in cases and controls according to phase of menstrual cycle (figure 1) suggested that concentrations in cases were higher than those in controls in the follicular phase and at mid-cycle, but not in the late luteal phase, and similarly the subgroup analyses of breast cancer risk showed larger ORs in the follicular phase and at mid-cycle than in the luteal phase, but these differences were not statistically significant.

Concentrations of all nine hormones, apart from androstenedione, were associated with BMI. Total oestradiol was inversely associated with BMI, whereas free oestradiol was positively associated with BMI because of the strong inverse association of SHBG with BMI. Interpretation of these observations is difficult, but if free oestradiol is a reliable index of bioavailable oestradiol then obese premenopausal women are exposed to a slightly more oestrogenic environment. Oestrone was also positively associated with BMI, perhaps because of increased peripheral aromatisation of androstenedione, as in postmenopausal women.[24] Progesterone concentrations were lower in obese than in non-obese women, whereas DHEAS and testosterone were positively associated with BMI. Similar findings for oestrogens and progesterone have been reported among regularly menstruating women in the BioCycle Study,[25] and in massively obese premenopausal women.[26]

Parity was not strongly associated with any of the hormones, but showed an inverse association with calculated free testosterone. Some previous studies in younger premenopausal women have suggested that early menarche and nulliparity are associated with oestrogen levels,[27,28] but in the current study none of the hormones or SHBG were significantly associated with age at menarche, and none of the oestrogen measures was associated with parity.

Concentrations of androstenedione, DHEAS, testosterone, and free testosterone were higher in women who smoked the most cigarettes and drank the most alcohol than they were in respective groups of non-smokers and non-drinkers. Very similar associations were noted in postmenopausal women.[29] The mechanism might involve stimulation of hormone synthesis by the adrenal glands.[30]

Women who had previously used hormonal contraceptives had lower concentrations of oestradiol, oestrone, androstenedione, and SHBG. Whether these associations are causal is unclear, as is what mechanism could be involved, although it might involve long-term effects on the liver.[31]

Sex hormones might mediate the effects of some risk factors on the development of breast cancer. For example, the increase in risk of breast cancer caused by alcohol[32] might be attributable to increased serum concentrations of sex hormones, although it could also be attributable to other effects of alcohol. BMI is inversely associated with the risk of breast cancer in premenopausal women,[33] and this association might be related to the effects of obesity on hormone levels. We noted that total oestradiol was inversely related to BMI and positively associated with risk, which is compatible with the idea that the lower risk in obese women is attributable to lower oestradiol, but this interpretation is complicated by the fact that we noted that free oestradiol was positively associated with BMI, as were oestrone and the androgens DHEAS, testosterone, and free testosterone. Luteal phase progesterone was also lower in obese than normal weight women, perhaps because of a higher probability of anovulatory cycles in obese women;[34] our analyses do not show any association of progesterone with risk of breast cancer, but the reliability of progesterone measurements is low and more data are needed before conclusions can be made that progesterone is not a determinant of breast cancer risk.

The strengths of this analysis were that the data and serum samples were collected on average several years before diagnosis, that we included almost all the available data from published studies worldwide, and we were able to adjust for phase of cycle and for other potential risk factors. The total sample size was moderately large for most of the hormones, but the power was low for the subgroup analyses. Many statistical tests are reported, therefore some of the nominally significant results may be due to chance.

A potential limitation was that the study designs and methods for measurement of hormones and other risk factors were not standardised. For example, studies variably used forward or backward dating for determination of when blood was collected in the menstrual cycle, and, because of differences in progesterone measurement between studies, we were unable to distinguish ovulatory versus anovulatory cycles. Furthermore, hormone concentrations varied substantially between studies. Some of this variation in mean hormone concentrations between studies was due to differences in the timing of sample collection, for example the relatively low mean oestradiol concentrations in the follicular phase in NHS-II samples that were collected on days 3–5 of the cycle, and the relatively high mean luteal phase progesterone concentrations in NHS-II and ORDET samples that were collected in the middle of the luteal phase. Some of the variation between studies probably shows differences in assay methods. The accuracy of assay methods varies, and assays that incorporate an extraction step are more accurate than are direct non-extraction assays.[35] Ideally, assays would be standardised and use the most accurate methods available, but in the present analysis our aim was to make the best use of the data available. To allow for differences in absolute hormone concentrations between assay laboratories we used study-specific quintiles of hormone concentrations.[21] This approach assumes that the true concentrations across the quintiles are similar in all the studies, and if this assumption is not correct then the estimates of ORs might be biased. However, because heterogeneity in risk estimates was not evident between studies or between assay methods (extraction *vs* non-extraction) this assumption does seem reasonable. Random error in laboratory estimates would lead to some underestimation of the associations observed.

Another potential limitation was that we used diagnosis of breast cancer before age 50 years as a surrogate for diagnosis before menopause. The median age at menopause in high-income countries is typically older than 50 years;[36] therefore, with the cutoff of 50 years most women would have been premenopausal at the time of breast cancer diagnosis, and the few women who were postmenopausal at diagnosis would on average have had menopause recently. Furthermore, most women in these studies were of white European ethnic origin, and further data for women with other ethnic origins would be valuable.

This collaborative analysis noted a positive association between sex hormones and breast cancer risk in premenopausal women. Whether or not this association is causal is not known, but plausible biological mechanisms exist that could explain such an effect, such as an increase in the mitotic rate of breast epithelial cells leading to an increased risk of mutations and the stimulation of the growth of early tumours.[37] The magnitude of the reported association was modest, but the true association could be substantially larger because of measurement error in the assessment of long-term premenopausal hormone levels. More data will become available in the next few years, both from extended follow-up of some of the studies in the present collaborative analysis, and from some new large studies such as the Breakthrough Generations Study[38] and UK Biobank.[39] More robust estimates are needed of the overall associations and associations in subgroups, as well as further analyses to clarify the relative importance of the different sex hormones, and to determine the environmental and genetic factors that cause differences in hormone levels in premenopausal women.

Contributors

All named members of the collaborative group were authors of this report. TJK coordinated the collaborative group, drafted the manuscript, did the literature search, and contributed to study design and data interpretation. PNA centralised the data, did the statistical analyses, and contributed to interpreting the data. GKR and RCT contributed to study design, data interpretation, and writing. All coauthors from the collaborating studies contributed to data collection, data interpretation, and writing.

Endogenous Hormones and Breast Cancer Collaborative Group

Writing committee—T J Key, P N Appleby, G K Reeves, R C Travis (Cancer Epidemiology Unit, Nuffield Department of Population Health,

University of Oxford, Oxford, UK). *Coauthors from collaborating studies*—A J Alberg (Department of Epidemiology, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA; and Hollings Cancer Center, Medical University of South Carolina, Charleston, SC, USA), A Barricarte (Navarre Public Health Institute, Pamplona, Spain; Consortium for Biomedical Research in Epidemiology and Public Health [CIBER Epidemiología y Salud Pública-CIBERESP], Spain), F Berrino, V Krogh, S Sieri (Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy), L A Brinton (National Cancer Institute, Bethesda, MD, USA), J F Dorgan (Fox Chase Cancer Center, Philadelphia, PA, USA), L Dossus (INSERM, Centre for Research in Epidemiology and Population Health [CESP], Nutrition, Hormones and Women's Health team, Villejuif, France; University Paris Sud, Villejuif, France; and IGR, Villejuif, France), M Dowsett (Institute of Cancer Research, London, UK), A H Eliassen (Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; and Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA), R T Fortner (Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; and Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA), S E Hankinson, for the Nurses' Health Study II research group (Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA; and Division of Biostatistics and Epidemiology, School of Public Health and Health ciences, University of Massachusetts, Amherst, MA, USA), K J Helzlsouer (Department of Epidemiology, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA; and Prevention and Research Center, Mercy Medical Center, Baltimore, MD, USA), J Hoffman-Bolton (Department of Epidemiology, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA; and George W Comstock Center for Public Health Research and Prevention, Washington County, MD, USA), R Kaaks (Deutsches Krebsforschungszentrum, Heidelberg, Germany), L L Kahle (Information Management Services, Rockville, MD, USA), K Koenig and A Zeleniuch-Jacquotte (Department of Environmental Medicine, New York University School of Medicine, New York, NY, USA), P Muti (McMaster University, Hamilton, Canada), K Overvad (Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark), P H M Peeters (Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands; and School of Public Health, Faculty of Medicine, Imperial College, London, UK), E Riboli (School of Public Health, Imperial College, London, UK), S Rinaldi (International Agency for Research on Cancer, Lyon, France), D E Rollison (Department of Interdisciplinary Oncology, H Lee Moffitt Cancer Center, Tampa, FL, USA), F Z Stanczyk (University of Southern California Keck School of Medicine, Los Angeles, CA, USA), D Trichopoulos (Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA; Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece; and Hellenic Health Foundation, Athens, Greece), S S Tworoger (Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA), P Vineis (HuGeF Foundation, Torino, Italy; and Imperial College, London, UK).

Conflicts of interest

All members of the writing committee declare that they have no conflicts of interest.

References

1 MacMahon B, Cole P, Brown J. Etiology of human breast cancer: a review. *J Natl Cancer Inst* 1973; **50:** 21–42.

2 Endogenous Hormones and Breast Cancer Collaborative Group. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst* 2002; **94:** 606–16.

3 Kaaks R, Rinaldi S, Key TJ, et al. Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition. *Endocr Relat Cancer* 2005; **12:** 1071–82.

4 Zhang X, Tworoger SS, Eliassen AH, Hankinson SE. Postmenopausal plasma sex hormone levels and breast cancer risk over 20 years of follow-up. *Breast Cancer Res Treat* 2013; **137:** 883–92.

5 Helzlsouer KJ, Gordon GB, Alberg AJ, Bush TL, Comstock GW. Relationship of prediagnostic serum levels of dehydroepiandrosterone and dehydroepiandrosterone sulfate to the risk of developing premenopausal breast cancer. *Cancer Res* 1992; **52:** 1–4.

6 Helzlsouer KJ, Alberg AJ, Bush TL, Longcope C, Gordon GB, Comstock GW. A prospective study of endogenous hormones and breast cancer. *Cancer Detect Prev* 1994; **18:** 79–85.

7 Dorgan JF, Stanczyk FZ, Kahle LL, Brinton LA. Prospective case-control study of premenopausal serum estradiol and testosterone levels and breast cancer risk. *Breast Cancer Res* 2010; **12:** R98.

8 Kaaks R, Berrino F, Key T, et al. Serum sex steroids in premenopausal women and breast cancer risk within the European Prospective Investigation into Cancer and Nutrition (EPIC). *J Natl Cancer Inst* 2005; **97:** 755–65.

9 Thomas HV, Key TJ, Allen DS, et al. A prospective study of endogenous serum hormone concentrations and breast cancer risk in premenopausal women on the island of Guernsey. *Br J Cancer* 1997; **75:** 1075–79.

10 Eliassen AH, Missmer SA, Tworoger SS, et al. Endogenous steroid hormone concentrations and risk of breast cancer among premenopausal women. *J Natl Cancer Inst* 2006; **98:** 1406–15.

11 Tworoger SS, Missmer SA, Eliassen AH, et al. The association of plasma DHEA and DHEA sulfate with breast cancer risk in predominantly premenopausal women. *Cancer Epidemiol Biomarkers Prev* 2006; **15:** 967–71.

12 Fortner RT, Eliassen AH, Spiegelman D, Willett WC, Barbieri RL, Hankinson SE. Premenopausal endogenous steroid hormones and breast cancer risk: results from the Nurses' Health Study II. *Breast Cancer Res* 2013; **15:** R19.

13 Zeleniuch-Jacquotte A, Afanasyeva Y, Kaaks R, et al. Premenopausal serum androgens and breast cancer risk: a nested case-control study. *Breast Cancer Res* 2012; **14:** R32.

14 Micheli A, Muti P, Secreto G, et al. Endogenous sex hormones and subsequent breast cancer in premenopausal women. *Int J Cancer* 2004; **112:** 312–18.

15 Baglietto L, Severi G, English DR, et al. Circulating steroid hormone levels and risk of breast cancer for postmenopausal women. *Cancer Epidemiol Biomarkers Prev* 2010; **19:** 492–502.

16 Kabuto M, Akiba S, Stevens RG, Neriishi K, Land CE. A prospective study of estradiol and breast cancer in Japanese women. *Cancer Epidemiol Biomarkers Prev* 2000; **9:** 575–79.

17 Grant EJ, Neriishi K, Cologne J, et al. Associations of ionizing radiation and breast cancer-related serum hormone and growth factor levels in cancer-free female A-bomb survivors. *Radiat Res* 2011; **176:** 678–87.

18 Wysowski DK, Comstock GW, Helsing KJ, Lau HL. Sex hormone levels in serum in relation to the development of breast cancer. *Am J Epidemiol* 1987; **125:** 791–99.

19 Södergård R, Bäckström T, Shanbhag V, Carstensen H. Calculation of free and bound fractions of testosterone and estradiol-17 beta to human plasma proteins at body temperature. *J Steroid Biochem* 1982; **16:** 801–10.

20 Endogenous Hormones and Breast Cancer Collaborative Group. Free estradiol and breast cancer risk in postmenopausal women: comparison of measured and calculated values. *Cancer Epidemiol Biomarkers Prev* 2003; **12:** 1457–61.

21 Key TJ, Appleby PN, Allen NE, Reeves GK. Pooling biomarker data from different studies of disease risk, with a focus on endogenous hormones. *Cancer Epidemiol Biomarkers Prev* 2010; **19:** 960–65.

22 Muti P, Trevisan M, Micheli A, et al. Reliability of serum hormones in premenopausal and postmenopausal women over a one-year period. *Cancer Epidemiol Biomarkers Prev* 1996; **5:** 917–22.

23 Missmer SA, Spiegelman D, Bertone-Johnson ER, Barbieri RL, Pollak MN, Hankinson SE. Reproducibility of plasma steroid hormones, prolactin, and insulin-like growth factor levels among premenopausal women over a 2- to 3-year period. *Cancer Epidemiol Biomarkers Prev* 2006; **15:** 972–78.

24 Edman CD, MacDonald PC. Effect of obesity on conversion of plasma androstenedione to estrone in ovulatory and anovulatory young women. *Am J Obstet Gynecol* 1978; **130:** 456–61.

25 Yeung EH, Zhang C, Albert PS, et al. Adiposity and sex hormones across the menstrual cycle: the BioCycle Study. *Int J Obes (Lond)* 2013; **37:** 237–43.

26 Kopelman PG, Pilkington TR, White N, Jeffcoate SL. Abnormal sex steroid secretion and binding in massively obese women. *Clin Endocrinol (Oxf)* 1980; **12:** 363–69.

27 Trichopoulos D, Cole P, Brown JB, Goldman MB, MacMahon B. Estrogen profiles of primiparous and nulliparous women in Athens, Greece. *J Natl Cancer Inst* 1980; **65:** 43–46.

28 Apter D, Reinilä M, Vihko R. Some endocrine characteristics of early menarche, a risk factor for breast cancer, are preserved into adulthood. *Int J Cancer* 1989; **44:** 783–87.

29 Endogenous Hormones and Breast Cancer Collaborative Group. Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. *Br J Cancer* 2011; **105:** 709–22.

30 Mello NK. Hormones, nicotine, and cocaine: clinical studies. *Horm Behav* 2010; **58:** 57–71.

31 Johnson PJ. Sex hormones and the liver. *Clin Sci (Lond)* 1984; **66:** 369–76.

32 Collaborative Group on Hormonal Factors in Breast Cancer. Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58 515 women with breast cancer and 95,067 women without the disease. *Br J Cancer* 2002; **87:** 1234–45.

33 van den Brandt PA, Spiegelman D, Yaun SS, et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. *Am J Epidemiol* 2000; **152:** 514–27.

34 Grodstein F, Goldman MB, Cramer DW. Body mass index and ovulatory infertility. *Epidemiology* 1994; **5:** 247–50.

35 Stanczyk FZ, Lee JS, Santen RJ. Standardization of steroid hormone assays: why, how, and when? *Cancer Epidemiol Biomarkers Prev* 2007; **16:** 1713–19.

36 Stanford JL, Hartge P, Brinton LA, Hoover RN, Brookmeyer R. Factors influencing the age at natural menopause. *J Chronic Dis* 1987; **40:** 995–1002.

37 Key TJ, Pike MC. The role of oestrogens and progestagens in the epidemiology and prevention of breast cancer. *Eur J Cancer Clin Oncol* 1988; **24:** 29–43.

38 Swerdlow AJ, Jones ME, Schoemaker MJ, et al. The Breakthrough Generations Study: design of a long-term UK cohort study to investigate breast cancer aetiology. *Br J Cancer* 2011; **105:** 911–17.

39 Collins R. What makes UK Biobank special? *Lancet* 2012; **379:** 1173–74.