

Protocol to assess risk of breast cancer associated with use of hormone replacement therapy in real world settings: two nested case-control studies in primary care

Yana Vinogradova, Carol Coupland, Julia Hippisley-Cox

I. ABSTRACT

Hormone replacement therapy (HRT) is used to help women suffering from menopausal symptoms and reduce risk of osteoporosis. Although effective, the therapy may lead to an increased risk of breast cancer. Previous studies have not been powerful enough to investigate the risks associated with different types of HRT. The proposed nested case-control study aims to fill this gap. Women diagnosed with breast cancer between 1998 and 2018 will be matched to 5 controls without breast cancer by age, practice and calendar year. Excluding prescriptions in the year before the index date (date of diagnosis of breast cancer in the case), exposure to each type of HRT will be defined as at least one prescription for that HRT prior to the index date. Conditional logistic regression will be used to assess the risks associated with the different types of oestrogen and progestogen. The associations with duration, length of any gap since earlier use and method of application will also be analysed for the most common types of HRT.

Index Terms: hormone replacement therapy; breast cancer; epidemiology

II. INTRODUCTION

A. *What is the problem being addressed?*

Menopause is a natural stage of reproductive life in women, which usually occurs between 45 and 55 years. Women's oestrogen levels decline and menstrual periods become irregular and gradually disappear. Many women also experience other symptoms such as hot flushes, night sweats, mood changes, memory loss, vaginal dryness, a lack of interest in sex, headaches, and joint stiffness. For some women, quality of life may be severely affected. Prolonged lack of oestrogen also increases risk of osteoporosis.

In November 2015, the National Institute for Health and Care Excellence (NICE) published its first ever guidance on the menopause¹ and further recommendations from the International Menopause Society followed². A central theme in both guidelines is the need to provide patients with information on the short and longer-term risks and benefits of HRT, to help women make informed choices about which treatment, if any, to use to relieve menopausal symptoms.

Between 2002 and 2005, and following the publication of two large studies raising concerns about the safety profile of HRT, the use of HRT had halved. These were the US Women's Health Initiative in 2002³ and the UK Million Women study in 2003⁴, both based on a 10 to 15 year study period and mean follow up period of between 4 and 5 years. The new NICE guideline is likely to result in a resurgence of the use of HRT among women once it is disseminated. Media reports about HRT have also not always been accurate, so responding to national and international concerns about the current inadequacy of risk-benefit information by providing women and their doctors with robust information on the risk of breast cancer associated with different types of HRT is important and timely.

A specific research question highlighted in the NICE guidance is how the different HRT preparations affect risk of breast cancer. This is the area which our proposal will focus on through analysis of longitudinal data over a 20-year period. The results will improve the evidence base for HRT and help patients and healthcare professionals make treatment choices better tailored to the individual.

B. *Why is the research important in terms of improving health?*

Breast cancer is the most common cancer in women in the UK, with over 55,000 cases of invasive breast cancer diagnosed during 2015⁵. The peak incidence is among women aged 50-59 years, which is likely to reflect both the epidemiology of the disease and the availability of breast screening, now offered to all women aged 50-65 years. The average age at onset of menopause in the UK is 51 years. This varies widely, however, and 1 in 100 women experience premature ovarian insufficiency (menopause occurring before the age of 40 years). Most women – 8 out of 10 – experience some menopausal symptoms, typically lasting about 4 years after the last menstrual period but in about 10% of women continuing for up to 12 years⁶.

Following the 2015 NICE guidance on the use of HRT, more women are likely to consult their GP about the risks and

Submission Date 11 July 2019.

Yana Vinogradova is Research fellow at the University of Nottingham (email yana.vinogradova@nottingham.ac.uk).

Carol Coupland is Professor of Medical Statistics in Primary Care at the University of Nottingham (email: carol.coupland@nottingham.ac.uk).

Julia Hippisley-Cox is Professor of Clinical Epidemiology & General Practice at the University of Oxford (email: Julia.hippisley-cox@phc.ox.ac.uk).

benefits of HRT. The guidance distinguishes different age groups, such as those under the age of 40 who have had a premature menopause due to premature ovarian insufficiency, those undergoing the menopause as a result of medical or surgical treatment (including women with cancer) and older women experiencing the menopause naturally.

Fear of breast cancer deters many women from taking HRT even if they have debilitating menopausal symptoms⁷. It is essential, therefore, to understand which types of HRT are safest, not only for individual women, but also at the population level. Small increases in risk of breast cancer are likely to have a significant impact at a population level given the frequency of breast cancer and the numbers of women eligible for treatment.

There are many types of HRT available. These can vary by: the regimen (unopposed oestrogen, combined cyclical or combined continuous); the type of oestrogen (conjugated equine oestrogen or oestradiol) or progestogen (medroxyprogesterone acetate or norethisterone/norgestrel or dydrogesterone); the dose and duration of treatment; and the route of delivery (oral or transdermal).

There are, however, significant gaps in our knowledge regarding the risk of breast cancer with different types of HRT and different methods of administration. Better safety profile information is essential given the large number of women who may now be considered for HRT. Large primary care research databases, such as QResearch and CPRD, have detailed information on diagnoses and prescriptions linked to hospital, cancer registration and mortality records. Since these databases have more than 20 years of detailed prescription data for millions of women and tens of millions of HRT prescriptions, which are all linked to hard clinical outcomes, they provide an efficient resource for quantifying risks associated with the different HRT treatments.

Improved information for women on both the relative and absolute risks of breast cancer associated with different HRT types will allow women to make better informed decisions.

C. How does the existing literature support this study?

Evidence for risk of breast cancer among women taking HRT has been mixed. In a reanalysis of early studies, data from 51 studies covering 21 countries were pulled together and reanalyzed to assess the risk of breast cancer associated with use of HRT⁸. The analysis was based on 53,865 women in whom 5,482 were exposed to HRT and developed breast cancer. Without distinguishing the types of hormones, this study reported a 2.3% (95% CI 1.1% to 3.6%) increased risk associated with each year of HRT use.

The Women's Health Initiative placebo-controlled trial (WHI)³ treated 16,608 post-menopausal women with either a combined HRT comprising conjugated equine oestrogen and medroxyprogesterone acetate or a placebo for, on average, 5.3 years. They reported 166 women receiving the HRT and 124 women receiving the placebo who developed invasive breast cancer, showing an increased risk of 26% (hazard ratio 1.26, 95% CI 1.00 to 1.59)³ associated with the HRT treatment. A

second arm of this trial, which compared conjugated equine oestrogen with placebo in 10,739 women who had had a hysterectomy, identified 94 breast cancer cases in the treatment group and 124 in the placebo group, reporting a borderline 23% reduction (hazard ratio 0.77, 95% CI 0.59 to 1.01) in breast cancer risk⁹.

In a meta-analysis including the WHI and similar placebo-controlled, trials, the other studies were much smaller, contributing 130 extra breast cancer cases in patients exposed to combined HRT and 21 extra cases in those exposed to oestrogen only therapy to the WHI results¹⁰. This study reported very similar results to the WHI study – an associated 27% increased breast cancer risk associated with overall use of combined therapy (relative risk 1.27, 95% CI 1.03 to 1.56) and a borderline 21% decreased breast cancer risk associated with overall use of oestrogen only therapy (0.79, 95% CI 0.61 to 1.01).

A major observational study, the Million Women cohort study, recruited menopausal women at the time of mammography and assessed their exposure to HRT using a questionnaire. This study followed up women for, on average, 2.6 years and reported 1934 women who developed breast cancer on combined therapy with an associated doubled risk of breast cancer (relative risk 2.0, 95% CI 1.91 to 2.09). It also reported 991 women who developed breast cancer on oestrogen-only therapy with an associated 30% increased risk (relative risk 1.30, 95% CI 1.21 to 1.40)⁴. This, together with adverse effects on risk of venous thromboembolism and cardiovascular disease, led many to the conclusion that the risks of HRT outweighed the benefits and caused a significant drop in HRT use¹¹.

Many studies looked at the risk of breast cancer after discontinuation of HRT but the definition of past use varied between studies and the number of HRT users was not sufficient to draw any definite conclusions. The WHI study and the Million Women study reported that incidence rates of breast cancer rapidly declined after discontinuation but remained increased in the first two years^{12 13}. Collaborative reanalysis of 51 studies found that the risk of breast cancer became similar to never users 5 or more years after discontinuation⁸.

Other observational studies have been small, only assessing breast cancer risk associated with overall use of any HRT¹⁴⁻¹⁷ because they were not powered to distinguish between treatments or types of hormones. There is currently a lack of evidence about comparative risks of breast cancer between different types of HRT – both for different types of oestrogen (conjugated vs estradiol) and different progestogens.

Evidence is not available from the large RCTs of HRT³ because they all studied only conjugated equine oestrogen and medroxyprogesterone acetate. New randomised controlled trials are unlikely given the cost, timescale and ethical difficulties in designing a trial to look for adverse outcomes such as breast cancer. New knowledge to help clarify the risks and benefits of different types of HRT will, therefore, need to come from large long-term observational studies.

D. 1.4 What is the research question?

The study aim is to quantify how different preparations of HRT affect the risk of breast cancer in order to determine which preparations have the best safety profile. Our objectives are:

1. to quantify risks of breast cancer associated with HRT by regimen (unopposed oestrogen; oestrogen combined with progestogen; tibolone); type of oestrogen (conjugated equine oestrogen, oestradiol) or progesterone (medroxyprogesterone acetate, dydrogesterone, drospirenone, norethisterone acetate and norgestrel/levonorgestrel); dose and duration of treatment; and route of delivery (oral, transdermal);
2. to determine whether these risks vary by age and body mass index in order to identify women most at risk of developing breast cancer when taking HRT.

III. METHODS

A. Design

We will undertake two nested case control studies with cases of breast cancer and matched controls using the QResearch (Version 43) and CPRD (GOLD, December 2018) databases. QResearch is a large validated database including the records of 35 million patients registered with approximately 1500 English practices, all linked at patient level to hospital episode statistics (HES) and Office for National Statistics (ONS) mortality data. CPRD also consists of data routinely collected from GP computer systems. It currently includes 16 million patients registered with 711 UK practices, of which 405 have patient data linked to HES and ONS mortality data.

There are strong similarities and some differences between the databases. Both comprise records made by general practitioners of their consultations with patients augmented by extra information from the practice, and both use the same coding system for diagnoses and symptoms – READ codes. The computer systems used are different – EMIS for QResearch and VISION for CPRD databases, and there are differences in the regional distribution of the practices and practice size. Because of the different computer systems, recording of ethnicity, family history¹⁸ and the evaluation of Townsend scores are different. Only 40% of CPRD VISION practices contributed in the last calendar year, 2018.

B. Population and selection of cases and controls

For each database, we will identify an open cohort of women aged from 50 to 79 years, registered during the study observation period from 1st January 1998 to 31st July 2018 for QResearch and from 1st January 1998 to 12th November 2018 for CPRD. Practices will be included if they have been using the relevant clinical system for at least 12 months. Women will be eligible for inclusion in the cohort only if they have been registered with the practice for at least three years to

ensure that the prescribing data are complete for this minimum period. The study entry will be the latest of: 1st January 1998; practice registration date plus one year; patient registration date plus three years; date of 40th birthday.

Cases will be women in the cohort who have a first diagnosis of breast cancer during the observation period recorded on either the GP, hospital or mortality record. For QResearch, cases will also be identified using linked Cancer Registry data. We will match each case diagnosed with breast cancer with up to 5 controls, who are alive and registered with the same practice at the time of this breast cancer diagnosis (the index date). Controls for each case will be matched by practice, age, and calendar time using incidence density sampling¹⁹. Each control will be allocated an index date, which will be the date of first diagnosis of breast cancer for the matched case from any of the linked data sources. Potential controls with fewer than three years of information before this index date will not be used. Women with a previous diagnosis of breast cancer or mastectomy identified from GP or hospital records – for controls, prior to study entry, for cases, prior to six months before the index date – will be excluded.

C. Outcome measure

The outcome is an incident diagnosis of breast cancer. Breast cancer cases will be identified based on diagnoses recorded on the GP record or the linked hospital, Cancer Registry (QResearch only) or mortality records using ICD 10 code lists validated for such use in previous studies of cancer.²⁰

D. Exposure

HRT exposure will be defined as any preparation containing oestrogen described in British National Formulary (BNF) Chapter 6.4.1 Sex hormones²¹. As a part of HRT treatment, some general practitioners in the UK prescribe progestogens which are also used for contraception and listed in BNF Chapters 7.3.1 and 7.3.2, and topical preparations listed in BNF Chapter 7.2.1. Two types of oestrogen (conjugated equine oestrogen and estradiol) and two types of progestogen (progesterone-and-analogues and testosterone-analogues) will be considered. Three types of progesterone-related (medroxyprogesterone acetate, dydrogesterone and drospirenone) and two types of testosterone-related progestogens (norethisterone acetate and norgestrel/levonorgestrel) are commonly prescribed in the UK. Because tibolone and raloxifene are sometimes used for treatment of menopausal symptoms, these drugs will be also included in the analysis.

We will extract all prescriptions for HRT in cases and controls since the early 1990s. Prescriptions for cases in the

last year before the index date might be related to treatment for early symptoms of breast cancer which are similar to menopausal symptoms. So these prescriptions will not be included in order to reduce possible protopathic bias. Similarly, prescriptions in the year before the index date will not be included in controls.

A woman will be considered as exposed to HRT from the first prescription containing an oestrogen. If a woman had only oestrogen containing preparations she will be considered as exposed to oestrogen-only HRT. If she had prescriptions for any progestogen either as a part of HRT preparation or a separate drug after she commenced HRT treatment she will be considered as exposed to combined HRT.

A woman can be exposed to a number of different oestrogens and progestogens during her overall period of treatment with HRT. We will consider these hormones to be individual exposures and, where feasible, analyse separately. We will also assess risks associated with overall exposures to HRT, to oestrogens and to progestogens. Exposures to oral and transdermal (patches, subcutaneous and gels) preparations, and to HRT creams and vaginal preparations, will be also included in the analysis and will be analysed separately.

The study will look at long term exposure as well as medium and short term effects. Using definitions from similar studies using electronic records, we will categorise HRT by: any use; type of drug prescribed; types of oestrogen and progestogen; routes of delivery; duration; dose; recency.^{4 8}
11 14

For oestrogen, the dose will be categorised into: low dose (≤ 0.625 mg for oral equine oestrogen, ≤ 1 mg for oral oestradiol, ≤ 50 micrograms for transdermal oestradiol); and high dose (> 0.625 mg for equine oestrogen, > 1 mg for oestradiol, > 50 micrograms for transdermal oestrogen).²² We will calculate the median dosages for each type of oestrogen during a patient's overall HRT exposure period and use these in separate analyses.

Duration of use will be assessed by calculating the number of days of exposure. If the gap between the end of one prescription and the start of next is 90 days or less, we will consider exposure as continuous^{23 24} and group such prescriptions in courses running from the beginning of the first prescription to the end of the last in the group. If the gap is longer than 90 days, we will consider the treatments as separate courses. Overall duration will be assessed by summing the separate course lengths. We will categorise duration into 5 groups: less than 1 year; 1 year up to 3 years; 3 years up to 5 years; 5 years up to 10 years; 10 years or more.

We will investigate recency of use by calculating the gap in years between the estimated date for last use of HRT and the index date, categorising it as follows: within 1 to 2 years before the index date; between 2 and up to 5 years; between 5 and up to 10 years; 10 years or more.

Women will be considered as current or recent users if their last estimated date of use of HRT falls into the last one to five years before the index date (since prescriptions in the year before the index date are not included), otherwise they will be considered as past users. The duration of use will be analysed separately for current/recent and past users.

We will then investigate separately for current users and past users the effects of durations of use categorised as above. We will also combine durations with recency of use in the following categories: short-term current users (use of less than 3 years); long-term current use (use of 3 years or more); short-term past users; long-term past users.

The reference category for all exposures will be no records of HRT use prior to one year before the index date.

E. Confounders

Potential confounders will be variables which are risk factors for breast cancer or indications for HRT. Demographic information will include self-assigned ethnic group (White or not recorded; Indian; Pakistani; Bangladeshi; other Asian; Caribbean; Black African; Chinese; Other)²⁵ and social deprivation as measured by Townsend score based on the patients' postcodes²⁶. Family history will cover breast cancer (including those identified with the BRCA1, BRCA2 or TP53 where recorded)²⁷⁻²⁹, ovarian, uterine or cervical cancers and osteoporosis or hip fracture.

Clinical history will be based on the latest relevant record at least one year before the index date and will include: alcohol consumption^{28 30} (non-drinker, ≤ 1 unit/day, 2-3 units/day, 4-6 units/day, 7+ units /day); body mass index³¹, smoking status (non-smoker, ex-smoker, light smoker (< 10 /day), moderate smoker (10-19/day), heavy smoker 20+ /day)³⁰; diabetes³²; bipolar disorder or schizophrenia³³; osteoporosis; benign breast disease (fibrocystic disease, intraductal papilloma, fibroadenoma)²⁹; hysterectomy-oophorectomy (based on GP and HES records); menopausal symptoms (hot flushes, menorrhagia, vaginitis); uptake of mammography (from GP and from HES data, where available). History of other cancers (blood cancer, cervical cancer, colorectal cancer, lung cancer, melanoma, ovarian cancer, uterine cancer)³³ will be included. Use or duration use for other drugs will include: aspirin³⁴; NSAIDs³⁵; hormonal contraceptives³⁶; tamoxifen.

Previous studies have reported that early menopause is associated with a lower risk of breast cancer and late

menopause with higher risk^{12,37}. Primary care data sources do not as a norm include information about the onset of menopause, so, to identify this we will use medical records related to the menopause and records of bilateral oophorectomy. Records of tests for follicle-stimulating hormone or luteinising hormone will not be considered because they are not recommended for diagnosing menopause⁷. Cases and controls with records of menopause or oophorectomy at age earlier than 45 years will be considered as early menopause. Cases and controls older than 55 years at the index date and with records of menopause at age higher than 55 years will be considered as late menopause. Indications of early and late menopause will be included into all analyses.

F. Statistical analysis

The two studies using QResearch and CPRD will be conducted in the same way, selecting the same confounders and running the similar procedures. All observations will be from general practices in the UK or linked records, be from the same time period, have similar exposures and have used similar methods for recording outcomes.

Conditional logistic regression will be used to estimate odds ratios with 95% confidence intervals for the HRT exposure variables. Unadjusted odds ratios and odds ratios adjusted for all confounders listed above will be reported. Adjusted odds ratios from the conditional regression analyses of the two datasets will be pooled using a fixed effect model with inverse variance weights³⁸. Where there is any heterogeneity we will pool the results using a random effect model. For statistically significant findings, excess risks and numbers needed to harm will be estimated by combining adjusted odds ratios with incidence rates in the underlying cohort³⁹.

Because BMI, smoking status and alcohol consumption may be important confounders but have non-negligible numbers of missing data, multiple imputation will be used to impute these⁴⁰⁻⁴³. Ten imputed datasets will be created. Index year, case/control status, age, years of records, potential confounders, and exposure to hormonal replacement therapy and other drugs, will be included in the imputation model. Prior to inclusion in the model, the distribution for BMI will be tested and, if not normal, a transformation of values will be carried out. Characteristics of women with missing values and with complete data will be compared to check that it is plausible that the data are missing at random.

We will calculate the numbers needed to harm and excess breast cancer risk over one year in women who used different estrogens and different progestogens for more than three years and had their last prescription within the last 5 years. The rate of breast cancer for the unexposed population will be assessed

using the CPRD cohort where the women will be followed until their first HRT prescription or leaving the study whichever happens first.

G. Sub-group and sensitivity analyses

Subgroup analyses will be run by age group (50 to 59, 60 to 69, 70 to 79) and by BMI categories (normal up to 25kg/m², overweight more than 25kg/m² and up to 30kg/m², obese 30kg/m² or more). When identifying BMI groups, only recorded BMI values will be used.

Sensitivity analyses will be run addressing different assumptions. In QResearch, all patients are linked to HES and ONS data and to cancer registry data, and almost all have a valid patient-level Townsend deprivation score. In CPRD, only 60% of practices are linked and have a valid patient-level Townsend score. For CPRD, we will run the analyses based on all available practices without adjusting for Townsend score, but the analyses will then be repeated on the subgroup of patients linked to HES and ONS data. We will also use patient-level Townsend deprivation index as a confounder for these analyses. If the results of these sensitivity analyses are different from the main analyses, we will publish them as the main findings and use them in the meta-analysis described above.

For the main analysis, we will use all HRT prescriptions recorded and adjust for the number of years of data. Some women, however, might have had HRT treatment before they joined the practice but will have been classified as unexposed for this study observation period. To address this misclassification bias we will run a sensitivity analysis for patients who have at least ten years of available data and will include all HRT prescriptions issued between 12 and 120 months before the index date.

For the main analysis, we will consider women who had HRT treatment within the last 1 to 5 years before the index date as current/recent users. Because the risk of breast cancer rapidly decreases after discontinuation we may be mixing lower risk in women who discontinued HRT in this interval with higher risk in continuing users. To address this, we will run a sensitivity analysis defining women who had HRT in the last 1 to 2 years as current users and women who stopped earlier as past users.

For the main analysis, we will consider exposure as continuous if the inter-prescription gaps are shorter than 90 days. There is, however, no information about when exactly the drugs were taken. It is possible that some women had an unaccounted supply of tablets or unused tablets from a previous prescription, so altering the actual inter-prescription gaps and affecting the continuity of exposure. To address this possible underexposure, we will run a sensitivity analysis

where the exposure is defined as the period of time between the start of the first prescription to the estimated end of the last prescription (or the end of follow-up, whichever occurs first) irrespective of gaps.

Regarding the assumption of missing at random in multiple imputation, a sensitivity analysis will be run restricted to women without missing data for BMI, smoking status or alcohol consumption. Regarding the assumption of categorizing people with missing values for ethnicity as White, a sensitivity analysis including an extra category of unrecorded ethnicity will also be done.

A 1% level of statistical significance will be used to allow for the multiple comparisons. Results will be presented as odds ratios and 95% confidence intervals to facilitate comparisons with other studies. Stata v 15 will be used for all the analyses.

IV. STRENGTHS AND LIMITATIONS

The main strength of the study is its generalisability and the large scale. Designing a two-database study will not only allow us to provide more precise estimates but will also increase the statistical power, facilitating investigation of less common exposures. Such a study will be able not only to investigate risks associated with different hormones used in HRT preparations, but also different durations of exposures to the hormones. There will also be sufficient observations to separate current from past users in the analyses.

The study will limit recall bias because exposure is assessed using prescription information. Using linked data as well as GP records for identifying cases will also reduce ascertainment bias for the outcome.

The limitations of the study will include uncertainty about the date of menopause, parity and breast feeding. These variables are not consistently recorded over the duration of the study so will have to be omitted. According to earlier research, HRT treatment started in the first 5 years after the menopause is associated with higher breast cancer risk than a later start^{12,44}. To address this limitation, we will use all available information to identify women with early and late menopause and also will run a subgroup analysis for different age categories.

Another limitation is the potential misclassification of exposure to HRT. We do not know with certainty whether a woman has filled a prescription or whether/when she started taking a prescribed medication. We do not see, however, any reason why this should differ between cases and controls. These two potential sources of misclassification are likely to be small, but might both shift odds ratios towards unity.

V. SUMMARY OF PATIENT ENGAGEMENT PLANS

The research questions have been identified by the NICE guideline group on HRT, which was published in 2015 and as such have been subject to public consultation at all stages of the development and publication of the guideline in accordance with NICE's policy and procedures. In addition, one of the authors (JHC) has discussed the proposal informally with a number of peri-menopausal and postmenopausal women, including several general practitioners.

As part of our research we will also involve a group of women who have used hormone replacement therapy, in order to discuss: their experiences of the treatments; any problems they experienced; whether this affected their ability or willingness to maintain treatment; and their concerns about risks. We will describe to them our proposed research and ensure that any issues they might raise are considered and appropriately incorporated. We also intend to disseminate our findings through general practices and in community settings.

A. Lay summary

At a certain age, or sometimes after specific health problems, women stop having periods and enter a life stage called the menopause. This is characterised by low levels of certain hormones, and during this period many women experience unpleasant symptoms, such as hot flushes, sweats or depression. They are also at increased risk of developing bone fragility, heart disease and urinary problems. Hormone replacement therapy (HRT) was introduced to relieve unpleasant symptoms and reduce risk of chronic health problems. There are different therapy types, which depend on the symptoms experienced by an individual. Some women require a treatment based only on the oestrogen hormone, others may need a combination of oestrogen and another hormone called progesterone. These therapies can also be administered in different ways – as tablets, patches or a cream.

Although all these treatments are effective in managing menopausal symptoms, they could lead to an increased risk of breast cancer. A recently issued guideline from the National Institute for Health and Care Excellence (NICE) has stressed that research findings from studies trying to estimate the risk of developing breast cancer as a result of taking HRT are still not clear, so not a good basis for decision-making by doctors or patients. In particular, studies so far have not been able to quantify different levels of risk for different treatments, regimens or application methods. This is because they had access to only a relatively small number of patients

for each treatment, making comparisons of outcomes unreliable.

This study will investigate risks of breast cancer from all types of hormone replacement therapy over a 20-year period from 1998 to 2018. We shall use two large databases, containing records from over 1700 English general practices and their associated hospital patient records. Patient confidentiality will be absolute because the information in these databases has been anonymised. We will compare the HRT treatment prescription records of all women who developed breast cancer with those of women who did not. We will take into account other health conditions and patient characteristics which might affect the risk of breast cancer to ensure that our results properly demonstrate the effects of the different treatments rather than other factors.

For single-hormone and for hormone-combination therapies we will investigate which specific types of included hormones have the lowest associations with risk of breast cancer, and will investigate risks related to different ways of administering the drugs (tablets, patches or creams).

The findings will provide much clearer and more detailed information for women and their doctors about breast cancer risks related to different types of HRT to help them make the most appropriate treatment choice.

VI. OTHER INFORMATION

A. Acknowledgements

The authors thank patients and EMIS practices who contribute to the QResearch® database, and EMIS and the Universities of Nottingham and Oxford for expertise in establishing, developing and supporting the QResearch database. We also thank patients and VISION practices who contribute to CPRD, and the NIHR and the MHRA for funding it. We thank the Division of Primary Care, University of Nottingham for contributing to the CPRD license. We also acknowledge that HES data to be used in this analysis are re-used by permission from the NHS Digital, who retain the copyright. We thank the Office of National Statistics for providing the mortality data. ONS and NHS Digital bear no responsibility for the analysis or interpretation of the data. This project involves data derived from patient-level information collected by the NHS, as part of the care and support of cancer patients. The data is collated, maintained and quality assured by the National Cancer Registration and Analysis Service, which is part of Public Health England (PHE). Access to the data was facilitated by the PHE Office for Data Release.

B. Approvals

The project has been reviewed in accordance with the QResearch® agreement with NRES Committee East Midlands – Derby [reference 18/EM/0400]. The protocol for CPRD has been approved by The Independent Scientific Advisory Committee for MHRA Database Research (N 16_275R).

C. Competing Interests

JHC is professor of clinical epidemiology at the University of Oxford and co-director of QResearch® – a not-for-profit organisation which is a joint partnership between the University of Oxford and Egton Medical Information Systems (the leading commercial supplier of IT for 60% of general practices in the UK). JHC is also a paid director of ClinRisk Ltd which produces open and closed source software to ensure the reliable and updatable implementation of clinical risk equations within clinical computer systems to help improve patient care. CC is Professor of Medical Statistics at the University of Nottingham and a paid consultant statistician for ClinRisk Ltd. This work and any views expressed within it are solely those of the co-authors, and not of any affiliated bodies or organisations.

VII. REFERENCES

1. National Institute for Clinical Excellence. Menopause: diagnosis and management. NICE guideline NG23, 2015.
2. Baber RJ, Panay N, Fenton A. 2016 IMS Recommendations on women's midlife health and menopause hormone therapy. *Climacteric* 2016;19(2):109-50. doi: 10.3109/13697137.2015.1129166
3. Writing Group for the Women's Health Initiative Investigators. Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results From the Women's Health Initiative Randomized Controlled Trial. *JAMA* 2002;288(3):321-33.
4. Million Women Study Collaborators. Breast cancer and hormone replacement therapy in the Million Women Study. *Lancet* 2003;362:419-27.
5. Cancer Research UK. Risk factors for Breast Cancer 2015 [Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/risk-factors>].
6. National Collaborating Centre for Women's and Children's Health. Menopause. Full guideline. *National Institute for Clinical Excellence* 2015 (Version 1.5)
7. National Institute for Clinical Excellence. Menopause: diagnosis and management. *NICE guideline NG23* 2015
8. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. *Lancet* 1997;350:1047-59.
9. The Women's Health Initiative Steering Committee. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: The women's health initiative randomized controlled trial. *JAMA* 2004;291(14):1701-12.
10. Gartlehner G, Patel SV, Feltner C, et al. Hormone therapy for the primary prevention of chronic conditions in postmenopausal women: Evidence report and systematic review for the us preventive services task force. *JAMA* 2017;318:2234-49.
11. Parkin DM. 10. Cancers attributable to exposure to hormones in the UK in 2010. *Br J Cancer* 2011;105:S42.

12. Beral V, Reeves G, Bull D, et al. Breast Cancer Risk in Relation to the Interval Between Menopause and Starting Hormone Therapy. *J Natl Cancer Inst* 2011;103(4):296-305.
13. Chlebowski RT, Kuller LH, Prentice RL, et al. Breast cancer after use of estrogen plus progestin in postmenopausal women. *The New England journal of medicine* 2009;360(6):573-87. doi: 10.1056/NEJMoa0807684
14. Jones ME, Schoemaker MJ, Wright L, et al. Menopausal hormone therapy and breast cancer: what is the true size of the increased risk? *Br J Cancer* 2016;115(5):607-15.
15. Liu J-Y, Chen T-J, Hwang S-J. The Risk of Breast Cancer in Women Using Menopausal Hormone Replacement Therapy in Taiwan. *International Journal of Environmental Research and Public Health* 2016;13(5):482. doi: 10.3390/ijerph13050482
16. Ewertz L, Mellekjær A, Poulson A, et al. Hormone use for menopausal symptoms and risk of breast cancer. A Danish cohort study. *British Journal of Cancer* 2005;92:1293-97.
17. Pesch B, Ko Y, Brauch H, et al. Factors Modifying the Association Between Hormone-Replacement Therapy and Breast Cancer Risk. *European Journal of Epidemiology* 2005;20(8):699-711. doi: 10.1007/s10654-005-0032-0
18. Kontopantelis E, Stevens RJ, Helms PJ, et al. Spatial distribution of clinical computer systems in primary care in England in 2016 and implications for primary care electronic medical record databases: a cross-sectional population study. *BMJ Open* 2018;8(2)
19. Lubin JH, Gail MH. Biased Selection of Controls for Case-Control Analyses of Cohort Studies. *Biometrics* 1984;40(1):63-75.
20. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of non-gastrointestinal cancers: series of nested case-control studies using QResearch and CPRD data. *Br J Cancer* 2013;109:795-806.
21. Royal Pharmaceutical Society of Great Britain, British Medical Association. *British National Formulary 66*. London: Pharmaceutical Press 2013.
22. Sweetland S, Beral V, Balkwill A, et al. Venous thromboembolism risk in relation to use of different types of postmenopausal hormone therapy in a large prospective study. *J Thromb Haemost* 2012;10:2277-86.
23. Lemaitre RN, Weiss NS, Smith NL, et al. Esterified Estrogen and Conjugated Equine Estrogen and the Risk of Incident Myocardial Infarction and Stroke. *Arch Intern Med* 2006;166(4):399-404.
24. Daly E, Vessey MP, Hawkins MM, et al. Risk of venous thromboembolism in users of hormone replacement therapy. *Lancet* 1996;348:977-80.
25. Gathani T, Ali R, Balkwill A, et al. Ethnic differences in breast cancer incidence in England are due to differences in known risk factors for the disease: prospective study. *Br J Cancer* 2013;110:224.
26. Townsend P, Davidson N. *The Black report*. London: Penguin 1982.
27. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease. *Lancet* 2001;358(9291):1389-99.
28. Pharoah PDP, Day NE, Duffy S, et al. Family history and the risk of breast cancer: A systematic review and meta-analysis. *Int J Cancer* 1997;71(5):800-809.
29. Zhou W-B, Xue D-Q, Liu X-A, et al. The influence of family history and histological stratification on breast cancer risk in women with benign breast disease: a meta-analysis. *J Cancer Res Clin Oncol* 2011;137(7):1053-60.
30. 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(11):1234-45.
31. Endogenous Hormones Breast Cancer Collaborative Group. Body Mass Index, Serum Sex Hormones, and Breast Cancer Risk in Postmenopausal Women. *J Natl Cancer Inst* 2003;95(16):1218-26.
32. Larsson SC, Mantzoros CS, Wolk A. Diabetes mellitus and risk of breast cancer: A meta-analysis. *Int J Cancer* 2007;121(4):856-62.
33. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithms to estimate future risk of common cancers in men and women: prospective cohort study. *BMJ Open* 2015;5(3)
34. Luo T, Yan H-M, He P, et al. Aspirin use and breast cancer risk: a meta-analysis. *Breast Cancer Res Treatm* 2012;131(2):581-87.
35. Takkouche B, Regueira-Mendez C, Etminan M. Breast Cancer and Use of Nonsteroidal Anti-inflammatory Drugs: A Meta-analysis. *J Natl Cancer Inst* 2008;100(20):1439-47.
36. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet* 1996;347(9017):1713-27.
37. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *The Lancet Oncology* 2012;13(11):1141-51. doi: [https://doi.org/10.1016/S1470-2045\(12\)70425-4](https://doi.org/10.1016/S1470-2045(12)70425-4)
38. Borenstein M, Hedges L, Higgins J, et al. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Syn Meth* 2010;1:97-111.
39. Bjerre L, LeLorier J. Expressing the magnitude of adverse effects in case control studies: "the number needed to be treated for one additional patient to be harmed". *BMJ* 2000;320:503-06.
40. Schafer J, Graham J. Missing data: our view of the state of the art. *Psychological Methods* 2002;7:147-77.
41. The Academic Medicine Group. Academic Medicine: problems and solutions. *BMJ* 1989;298:573-79.
42. Steyerberg EW, van Veen M. Imputation is beneficial for handling missing data in predictive models. *J Clin Epidemiol* 2007;60(9):979.
43. Moons KGM, Donders RART, Stijnen T, et al. Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology* 2006;59(10):1092.
44. Prentice RL, Manson JE, Langer RD, et al. Benefits and Risks of Postmenopausal Hormone Therapy When It Is Initiated Soon After Menopause. *Am J Epidemiol* 2009;170(1):12-23.