

## **Development of a cancer risk prediction tool for use in the UK Primary Care and community settings.**

Artitaya Lophatananon<sup>1</sup>, Juliet Usher-Smith<sup>2</sup>, Jackie Campbell<sup>3</sup>, Joanne Warcaba<sup>4</sup>,  
Barbora Silarova<sup>5</sup>, Erika A. Waters<sup>6</sup>, Graham A. Colditz<sup>6</sup> and Kenneth R. Muir<sup>1</sup>

<sup>1</sup> Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK.

<sup>2</sup>The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge.UK.

<sup>3</sup>Institute of Health and Wellbeing, The University of Northampton, Northampton. UK.

<sup>4</sup>Moulton Surgery, Northampton Lane, Northampton. UK.

<sup>5</sup>MRC Epidemiology Unit, School of Clinical Medicine, University of Cambridge, Cambridge. UK.

<sup>6</sup> Division of Public Health Sciences Department of Surgery, Washington University School of Medicine, St. Louis, Missouri. USA.

**Running title:** Development of cancer risk prediction tool

**Keywords:** cancers, risk prediction, primary care, community

### **Financial support:**

A. Lophatananon, K. R. Muir, J. Usher-Smith, J. Campbell, J. Warcaba and B. Silarova are funded by an Innovation Grant from the Cancer Research UK – BUPA Foundation Fund (C55650/A20818).

A. Lophatananon and K.R. Muir are also funded by ICEP ("This work was also supported by CRUK [grant number C18281/A19169]")

J. Usher-Smith is also supported by the National Institute for Health Research Clinical Lectureship and B. Silarova was supported by the Medical Research Council [MC\_UU\_12015/4].

E. A. Waters, G. A. Colditz are funded in part by the Foundation for Barnes-Jewish Hospital, St Louis, MO.

**Correspondence to:**

Dr Artitaya Lophatananon, Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK M139PL

E-mail: [artitaya.lophananon@manchester.ac.uk](mailto:artitaya.lophananon@manchester.ac.uk)

## **Abstract.**

Several multivariable risk prediction models have been developed to assess an individual's risk of developing specific cancers. Such models can be used in a variety of settings for prevention, screening and guiding investigations and treatments. Models aimed at predicting future disease risk that contains lifestyle factors may be of particular use for targeting health promotion activities at an individual level. This type of cancer risk prediction is not yet available in the UK. We have adopted the approach used by the well-established U.S. derived "YourCancerRisk" model for use in the UK population which allow users to quantify their individual risk of developing individual cancers relative to the population average risk. The UK version of YourCancerRisk" computes 10 year cancer risk estimates for 11 cancers utilising UK figures for prevalence of risk factors and cancer incidence. Since the prevalence of risk factors and the incidence rates for cancer are different between the US and the UK population, this UK model provides more accurate estimates of risks for a UK population. Using an example of breast cancer and data from UK Biobank cohort we demonstrate that the individual risk factor estimates are similar for the US and UK populations. Assessment of the performance and validation of the multivariate model predictions based on a binary score confirm the model's applicability. The model can be used to estimate absolute and relative cancer risk for use in Primary Care and community settings and is being used in the community to guide lifestyle change.

## **Introduction.**

Over recent years, there has been a growth in the development of risk prediction models for cancer and other diseases (1-6). These models provide a range of estimates of future risk of developing disease for applications in prevention, screening, diagnosis and treatment. Most of them are disease specific. In general, risk algorithms include phenotypic information such as sex, age, and lifestyle factors. Some algorithms may also allow for the incorporation of emerging “omics” based factors and other biomarkers (7).

Few cancer risk prediction models have included the modifiable risk components such as physical activity, diet and smoking. One such model that has been developed for the US is the "YourCancerRisk" model which has been used as a tool for education as well as providing an approach to quantifying the effects of changing key lifestyle exposures.

This was subsequently expanded into "YourDiseaseRisk" (8) which extends the range of endpoints to include 12 of the commonest types of cancers in the USA and 6 other chronic diseases (e.g., chronic bronchitis, stroke, emphysema, heart disease, diabetes, and osteoporosis). The model was validated for ovarian, colon and pancreatic cancers in the U.S in Nurses' Health Study and the US Health Professionals cohorts. The results show the model to be well calibrated for ovarian and colon cancer in women and pancreatic cancer in men and moderately calibrated for colon cancer in men.

Discriminatory accuracy for pancreatic cancer showed a concordance index of 0.72, and for colon cancer in men and women concordance indices were 0.71, 0.67 respectively (9).

The “YourDiseaseRisk” model aims to predict the risks for individuals (aged 40 and above) of developing the 12 cancers relative to the general population. Uniquely, the approach adopted to develop such models involved extensive systematic reviews of existing studies and finding a consensus of expert opinions to identify risk factors and to summarise the level of evidence as “definite”, “probable” and “possible” causes of cancer. Risk points were then allocated according to the strength of the causal association and summed. Population average risk of cancer and cumulative 10-year risk were obtained from the US SEER data (10). Finally, individual ranking relative to the population average was determined.

The “YourDiseaseRisk” online tool has been available in the US since 2000. It is offered as an educational tool and in 2005, the site recorded 54 million hits with 6.2 million page views. It was first hosted at Harvard University and, in 2007, transitioned to the Siteman Cancer Center at Barnes-Jewish Hospital and Washington University School of Medicine (CITE Siteman/Wash U/ BJH only)(8).

The focus of this paper is to describe the steps taken to adapt the "YourDiseaseRisk" models focused on cancers for the UK population for use in Primary Care and community based settings. We also assess the utility of the approach by scoring the suggested risk factors in the UK Biobank cohort.

We have used the adapted UK version of the "YourDiseaseRisk" models (8) in a pilot study that used individual interviews to assess participants’ understanding and preferences for how such information is offered (publication in press). These results will be presented elsewhere.

## **Materials and Methods.**

### **Cancer Risk models development.**

We adapted the YourDiseaseRisk models for 11 cancers using the UK data. The 11 cancers chosen were lung, prostate, breast, kidney, bladder, colon, skin, stomach, pancreatic, uterine and ovarian cancer. Cervical cancer was not included as it required participants to disclose information on sexual history which was considered too sensitive.

In general terms, the information required to develop prediction models are: 1) the identified list of risk factors for inclusion; 2) point estimates of the relative risk for each risk factor; and 3) population prevalence for each of the exposures. To be able to compare individual risk to the population, further information such as cancer incidence by 10 year age bands is required.

### **Comparison of relative risk (RR) between the US and the UK.**

To illustrate the comparative RRs between the two populations, we used the UKBiobank national cohort and analysed the RRs for breast cancer. The UKBiobank female cohort consists of 273,467 women with age ranged between 40-69 years when recruited.

Participants were enrolled in the UK Biobank from April, 2007, to July, 2010, from 21 assessment centres across England, Wales, and Scotland using standardised procedures(11). The UK Biobank study was approved by the North West Multi-Centre Research Ethics Committee and all participants provided written informed consent to participate in the UK Biobank study. To date, the cohort has been followed up for 6 years. We computed RRs adjusted only for age. The results are presented in table 6.

### **Model validation-an example.**

To demonstrate model validation, we selected breast cancer risk prediction as an example. The UKBiobank cohort was used for the validation exercise. For breast cancer cases, we used ICD10, ICD9 and self-reported codes (verified by the UKBiobank health professionals) and only incident cases were included in the analysis. For controls, we used 2 comparison groups. Firstly only those subjects with no coded record in ICD10, ICD9 of cancer and secondly the coded with no cancer and no other self-reported illness. The total number of incident breast cancer cases was 3,378 and the total number of non-cancer controls was 235,603 or healthy controls was 59,731. We coded all variables (except Tamoxifen/Raloxifene usage as such data was not available) present in the breast cancer risk prediction model based on presence or absence of the exposure for each individual as illustrated in Table 3. To demonstrate the validity of the combined risk factors identified for the model we calculated the area under the curve of the model based on including all factors scored as a binary variable and generated calibration plots -the observed and expected proportions compare within the groups formed by the Hosmer-Lemeshow test. All analyses were performed using STATA 14 (12).

### **Results.**

To develop the UK version of the “YourDiseaseRisk” models we have assumed that the risk factors for these cancers in the US and UK are the same and have obtained the list of risk factors and point estimates of relative risk for each risk factor from the US “YourDiseaseRisk” models. Since the "YourDiseaseRisk" model was first developed more than a decade ago and to ensure we use the most updated version, we started the process by extracting risk factors for each cancer from the "YourDiseaseRisk" online

version(8) (Step 1). Once we created a list of risk factors for each cancer, we assigned point estimates of relative risk (RR) for each factor (Step 2). These point estimates of RR's were extracted from the original publication (13, 14), however if any factors did not have any cited value, we performed a literature search to obtain any missing estimates. To maintain consistency across RR values, we used figures from publications by the Colditz study group and those with a cohort study design (examples of references are depicted in Table 1). For example, as the multivitamin factor was not listed in the original article, we therefore chose and applied a RR of 0.7 from Zhang et al (15).

As the original report applied RRs from compiling evidence from the US cohort studies over a period of time, it is important to justify the use of the RRs published by Colditz et al(14) in the UK models. We demonstrated the similarities/variations of risks between the two populations (US and UK). To do this we have presented as an example the relative risks for breast cancer derived from the UKBiobank study (Table 2). The majority of the point estimates are similar and convert to the same risk score when using values in Table3. The two exceptions were for multivitamin use and physical activity where the protective effects for both of these factors were less pronounced in the UK as compared to the US estimate.

To obtain the population prevalence (step 3) for each of the exposures we then reviewed the literature on the UK prevalence of each factor for men and women for all 11 cancers. The criteria for publication selection included a) UK prevalence data from National surveys (16) or prevalence derived from large cohort studies representative of the general population in the UK, or b) if no data were available from those sources,



information/figures from cohort studies in European countries. As an example, the colon cancer risk factors, RR and references chosen for prevalence are shown in Table 1.

### **Computing risk scores.**

Once information on the UK prevalence of each risk factor was obtained, we then applied a score to each risk factor using the same scheme as presented in the original paper (summary as shown in Table 3).

This risk score is used to compute two further scores – the population average risk score and an individual risk score relative to the population average. The population average score is calculated by multiplying the risk score of each factor by the population prevalence of that particular factor. To prevent negative scores, we chose the direction of each risk factor to make the population average score the highest possible. Taking physical activity, for example, the prevalence of carrying out 3 or more hours of total leisure-time physical activity per week in the UK population is 23%. This figure means that 77% of the population do not do physical activity regularly at this particular level. When we apply a prevalence of 77% then the assigned score instead of being -10 (for those who are doing regular exercise) will be +10. This conversion allows us to demonstrate the change in the individual risk score following any change in factors that are modifiable. Summation of these scores produces the average population score.

The risk score for a given individual relative to the population average is based on presence or absence of each factor. An example is illustrated in Table 4. The summation of scores for each risk factor for an individual provides the total risk score for a particular person. That total risk score is then divided by the average population score to give an individual index which is relative to population score ( $35/27 = 1.3$ ). As

mentioned earlier, participants can see how their risk changes if they adopt suggested behaviours, for example by choosing to do regular exercise the total score for the individual illustrated in Table 4 can be reduced by 10 points (from 35 to 25), making their index score relative to the population reduced from 1.3 to 0.8. This calculation aims to illustrate to each individual the effect of a particular lifestyle or behaviour change leading to cancer risk reduction.

### **Conversion of individual index score to 5-category cancer level of risk.**

The index score for an individual can then be further transformed into a level of risk. A numeric factor indicative of the strength of the risk level is assigned to the individual index score (Table 5). This is done to give an average value of the range of individual index scores as a single numeric estimate that reflects the risk level.

For the individual in Table 3, for example, the individual index score is 1.3, which is equivalent to “above average risk” and gives a numeric factor of 1.5.

### **10 Year estimated cancer risk.**

To enable estimation of an individual’s 10 year estimated cancer risk we calculated the average 10 year estimated risk for different ages and sexes of the UK population for all 11 cancers. We used the “Current Probability” method proposed by Esteve et al in 1994 (17). This method uses a life-table approach for calculating the risk of developing cancer, and takes into account the likelihood of dying from other causes. The method also requires information on deaths from all causes for each age-group. This method provides estimate of the 10 year risk of cancer.

We obtained age- and sex-specific cancer incidence and mortality rates and numbers from Cancer Research UK (18) and age- and sex-specific data on all-cause mortality

from the Office of National Statistic (ONS) which are available online (19). The following shows the specific data used for the calculation:

1. The annual number of cancer deaths (cancer mortality for males and females); data from CRUK 2010-2012.
2. The annual number of (registered) cancer cases for males and females; data from CRUK 2009-2011.
3. The annual number of deaths (all-cause mortality for males and females); data from ONS 2011.
4. The size of the mid-year population for males and females; data from ONS 2011.

From this data we computed 10-year cancer risk for each cancer in males and females in 10 year age bands from birth to over 80 years.

#### **Individual estimate 10 year cancer risk.**

Assuming the individual in Table 4 is a 40-year-old woman, the average 10-year risk of breast cancer (age 40-49) is 1.45% in the next 10 years (Table 6). Multiplying this figure by the numeric factor of 1.5 (Table 4) will give a risk of 2.2% (above the population average) or approximately 1 case in 45 women.

#### **Example of Breast cancer risk prediction model performance and validation based on binary scoring of risk factors.**

The AUC for breast cancer risk prediction model based on utilising a binary score for each available risk factor in the UKBiobank dataset was 0.58 (95%C.I 0.57-0.60) for a comparison group of controls with no cancer and 0.64 (95%C.I 0.63-0.66) for a

comparison group of controls with no cancer or other illnesses. Model calibration curves for both comparisons also suggested both models calibrated well (Figure 1).

## **Discussion.**

The “YourDiseaseRisk” model has been developed using three key data including an estimate of relative risk of each risk factor, the prevalence figure of exposure in the population and the 10 year estimated cancer risk. These figures can be acquired through literature review, National data archives or other organisations that compile these data publically (20).

The “YourDiseaseRisk” site has been launched since 2000 and update/review of information is an ongoing process. The model is an educational tool and can be used for cancer prevention in clinics, community settings, or by individuals simply seeking information on their own. The popularity of “YourDiseaseRisk” is reflected by the large number of hits or page views.

We have adopted their methodology and applied it to build a UK version and to demonstrate the applicability of the approach for other populations as well. The main strength of the approach is the use of large population based studies to estimate the parameters needed to generate the model.

To demonstrate the relative validity of the approach we chose breast cancer as an exemplar. We have demonstrated that the quantitative estimates of the individual risk factors apply to current UK population using data from the UKBiobank study.

Furthermore we performed a model validation of the combined risk factors using a binary scoring system and the results suggests such a model is reasonably calibrated and has moderate discriminatory power (0.58 for a comparison group of controls with

no cancer and 0.64 (95%CI 0.63-0.66) for a comparison group of controls with no cancer or other illnesses). The prediction performance of a fully specified model using the point estimates of each risk factor rather than a binary value is likely to be higher as it will be derived from more precise estimates of their individual effect. Directly comparative data is not yet available in the UKBiobank for all parameters to undertake such analyses.

There are many breast cancer risk prediction models but only a few that contained epidemiological factors and most have similar predictive capabilities (21-30). The majority of these models are extended versions of the Gail model(23). They have been shown to have good calibration but moderate discrimination ranging from 0.56-0.89 with few having been assessed in external validation analysis (23, 28, 31, 32). Colditz et al, further reported an AUC of 0.64 (95%CI 0.62–0.66) in their extended validation study using data from the Nurses' Health Study (32). The validation exercise, based on breast cancer example presented here, supports the conclusion that the model can be used in the UK population when the prevalence of risk factors is substituted with the UK figures.

There are, however, potential limitations to the approach that need to be taken into consideration. Firstly, although we were systematic in our literature reviews to obtain data on the point estimate RRs for risk factors missing from the original published “YourDiseaseRisk” model and the prevalence of all risk factors in the UK population, the studies we have chosen may not provide the most accurate estimates. Secondly, we have assumed that the prevalence of each risk factor does not change with age, that the risk associated with each risk factor is the same across all ages and both sexes, and that the risk factors do not interact with each other. These assumptions were also made in

the development of the “YourDiseaseRisk” models but, as noted in that report (14), may result in misclassification of risk for exposures across large age ranges and underestimate possible synergistic effects of exposures such as alcohol and smoking. As for the “YourDiseaseRisk” models, therefore, these models should be considered as a guide for assessing an individual’s risk of cancer in the UK rather than a precise estimate.

Risk prediction models are widely used for many diseases. With the concept that some cancers can be preventable, the “Your Disease risk” model developed by Colditz et al (14) provides a good platform because their model is based on modifiable factors that have been scrutinised and carefully selected by a panel of experts. The approach has wide utility in allowing for the rapid development of models for educational purposes.

In this paper we have described how we have adapted that model to produce a UK version for 11 cancers using data from the UK population. We are now using the model in the community in the UK. Going forwards we will evaluate how the tool affects perceptions of risk, how to best present the risk, and how the public understands their individual risk. This information will then inform future studies in the community exploring the potential for the use of this model to promote lifestyle change. Finally, we will extend and further validate this risk prediction in the UKBiobank cohort as more precise individual level data and longer term follow up data become available.

### **Acknowledgements.**

This work was not possible without the funding from the CRUK- BUPA Foundation Fund (C55650/A20818), ICEP CRUK [grant number C18281/A19169]), the Medical Research Council [MC\_UU\_12015/4], MAHSC Cancer Domain [L20-2015] and the

Foundation for Barnes-Jewish Hospital, St Louis, MO. We would like to thank Ms Kawthar-Al-Ajmi, PhD student at the University of Manchester for her help with preparation the UKBiobank data. This research has been done using the UK Biobank resources application number 5791.

## References

1. Usher-Smith JA, Walter FM, Emery JD, Win AK, Griffin SJ. Risk Prediction Models for Colorectal Cancer: A Systematic Review. *Cancer Prev Res (Phila)*. 2016 Jan;9(1):13-26.
2. Echouffo-Tcheugui JB, Kengne AP. Risk models to predict chronic kidney disease and its progression: a systematic review. *PLoS Med*. 2012;9(11):e1001344.
3. Tangri N, Kitsios GD, Inker LA, Griffith J, Naimark DM, Walker S, et al. Risk prediction models for patients with chronic kidney disease: a systematic review. *Ann Intern Med*. 2013 Apr 16;158(8):596-603.
4. Engel C, Fischer C. Breast cancer risks and risk prediction models. *Breast Care (Basel)*. 2015 Feb;10(1):7-12.
5. Gray EP, Teare MD, Stevens J, Archer R. Risk Prediction Models for Lung Cancer: A Systematic Review. *Clin Lung Cancer*. 2015 Dec 1.
6. Usher-Smith J, Emery J, Hamilton W, Griffin SJ, Walter FM. Risk prediction tools for cancer in primary care. *Br J Cancer*. 2015 Dec 22;113(12):1645-50.
7. Bruzelius M, Bottai M, Sabater-Lleal M, Strawbridge RJ, Bergendal A, Silveira A, et al. Predicting venous thrombosis in women using a combination of genetic markers and clinical risk factors. *J Thromb Haemost*. Feb;13(2):219-27.
8. Siteman Cancer Center at Barnes-Jewish Hospital and Washington University School of Medicine. Your Disease Risk. <http://www.yourdiseaserisk.wustl.edu/YDRDefault.aspx?ScreenControl=YDRGeneral&ScreenName=YDRAbout>; 2013 [cited 2016 20 July].
9. Kim DJ, Rockhill B, Colditz GA. Validation of the Harvard Cancer Risk Index: a prediction tool for individual cancer risk. *J Clin Epidemiol*. 2004 Apr;57(4):332-40.
10. Cancer Statistics [database on the Internet]. [www.seer.cancer.gov](http://www.seer.cancer.gov). 2016 [cited 26/07/2016].
11. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015 Mar;12(3):e1001779.
12. StataCorp. Stata Statistical Software: Release 14 College Station. Texas: StataCorp LP; 2015.
13. Tomeo CA, Colditz GA, Willett WC, Giovannucci E, Platz E, Rockhill B, et al. Harvard Report on Cancer Prevention. Volume 3: prevention of colon cancer in the United States. *Cancer Causes Control*. 1999 Jun;10(3):167-80.
14. Colditz GA, Atwood KA, Emmons K, Monson RR, Willett WC, Trichopoulos D, et al. Harvard report on cancer prevention volume 4: Harvard Cancer Risk Index. Risk Index Working Group, Harvard Center for Cancer Prevention. *Cancer Causes Control*. 2000 Jul;11(6):477-88.
15. Zhang S, Hunter DJ, Hankinson SE, Giovannucci EL, Rosner BA, Colditz GA, et al. A prospective study of folate intake and the risk of breast cancer. *JAMA*. 1999 May 5;281(17):1632-7.
16. Health Survey for England, 2013. [data collection] [database on the Internet]. <http://dx.doi.org/10.5255/UKDA-SN-7649-1>. 2015.
17. Esteve J, Benhamou E, Raymond L. Statistical methods in cancer research. Volume IV. Descriptive epidemiology. *IARC Sci Publ*. 1994(128):1-302.
18. Cancer Statistics for the UK [database on the Internet]. <http://www.cancerresearchuk.org/health-professional/cancer-statistics>. 2016 [cited 26/07/2016].
19. All Cause mortality [database on the Internet]. <https://www.ons.gov.uk/>. 2016 [cited 26/07/2016].



20. National Cancer Institute. Data & Software for Researchers. 2017 [cited 2017 23/04]; Available from: <https://seer.cancer.gov/resources/>.
21. Banegas MP, Gail MH, LaCroix A, Thompson B, Martinez ME, Wactawski-Wende J, et al. Evaluating breast cancer risk projections for Hispanic women. *Breast Cancer Research and Treatment*. 2012;132(1):347-53.
22. Boyle P, Mezzetti M, La Vecchia C, Franceschi S, Decarli A, Robertson C. Contribution of three components to individual cancer risk predicting breast cancer risk in Italy. *European journal of cancer prevention*. 2004;13(3):183-91.
23. Gail MH, Costantino JP, Pee D, Bondy M, Newman L, Selvan M, et al. Projecting individualized absolute invasive breast cancer risk in African American women. *Journal of the National Cancer Institute*. 2007;99(23):1782-92.
24. Lee C, Lee JC, Park B, Bae J, Lim MH, Kang D, et al. Computational Discrimination of Breast Cancer for Korean Women Based on Epidemiologic Data Only. *Journal of Korean Medical Science*. 2015 07/15 01/02/received 04/09/accepted;30(8):1025-34.
25. Lee E-O, Ahn S-H, You C, Lee D-S, Han W, Choe K-J, et al. Determining the Main Risk Factors and High-risk Groups of Breast Cancer Using a Predictive Model for Breast Cancer Risk Assessment in South Korea. *Cancer nursing*. 2004;27(5):400-6.
26. Matsuno RK, Costantino JP, Ziegler RG, Anderson GL, Li H, Pee D, et al. Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. *Journal of the National Cancer Institute*. 2011.
27. Novotny J, Pecen L, Petruzalka L, Svobodnik A, Dusek L, Danes J, et al. Breast cancer risk assessment in the Czech female population—an adjustment of the original Gail model. *Breast Cancer Research and Treatment*. 2006;95(1):29-35.
28. Park B, Ma SH, Shin A, Chang M-C, Choi J-Y, Kim S, et al. Korean risk assessment model for breast cancer risk prediction. *PLoS One*. 2013;8(10):e76736.
29. Ueda K, Tsukuma H, Tanaka H, Ajiki W, Oshima A. Estimation of individualized probabilities of developing breast cancer for Japanese women. *Breast Cancer*. 2003;10(1):54-62.
30. Wang S, Ogundiran T, Ademola A, Olayiwola OA, Adeoye A, Adeniji-Sofoluwe A, et al. Abstract 2590: Development and validation of a breast cancer risk prediction model for black women: findings from the Nigerian breast cancer study. *Cancer Research*. [10.1158/1538-7445.AM2016-2590]. 2016;76(14 Supplement):2590.
31. Rosner B, Colditz GA. Nurses' health study: log-incidence mathematical model of breast cancer incidence. *Journal of the National Cancer Institute*. 1996;88(6):359-64.
32. Colditz GA, Rosner B. Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. *American Journal of Epidemiology*. 2000;152(10):950-64.

Table 1. Colon cancer risk factors, its prevalence and reference information

<b>Risk factor</b>	<b>Definition</b>	<b>Male prevalence</b>	<b>Female prevalence</b>	<b>Source:</b>	<b>Population:</b>
<b>1. Family history</b>	Brother, sister or parent had colon cancer (1 <sup>st</sup> degree relatives affected with colon cancer).	6.0%	7.0%	Sandhu MS, Luben R, Khaw KT (2001) Prevalence and family history of colorectal cancer: implications for screening. J Med Screen 8(2): 69-72.	30,353 participants aged 45-74 were recruited from GP between 1993 and 1997 as part of the East Anglian component of the European Prospective Investigation into Cancer (EPIC-Norfolk).
<b>2. Obesity</b>	BMI≥27 kg/m <sup>2</sup>	52.0%	49.0%	Health England survey 2013 (Micro data). The Health Survey for England series was designed to monitor trends in the nation's health, to estimate the proportion of people in England who have specified health conditions, and to estimate the prevalence of certain risk factors and combinations of risk factors associated with these conditions.	UK population with a total of 2,362 males and 2,810 females aged ≥45.
<b>3. Saturated fat</b>	milk or dairy products ≥3 serving/day	9.0%	12.0%	Findings from the National Adult Nutrition Survey: Dairy Intakes and Compliance with Food Pyramid Recommendations among Irish Adults Aged 65 years and over	Irish population (database contain details data on servings, we selected high intake as ≥3.5 servings/day).
<b>4. Alcohol</b>	More than 7 servings per week	59.0%	41.0%	Health England survey 2013 (Micro data).	UK population with a total of 2,362 males and 2,810 females aged ≥45. Variable used for the analysis: total unit of drinks per week.
<b>5. Vegetables</b>	3 or more servings per day	14.0%	14.0%	Health England survey 2013 (Micro data).	UK population with a total of 2,362 males and 2,810 females aged ≥45. Variable used for the analysis: total portion of vegetable per day (adults aged 45+).
<b>6. Height</b>	5 ft. 7 in or taller	10.0%	7.0%	Health England survey 2013 (Micro data).	UK population with a total of 2,362 males and 2,810 females aged ≥45. Variable used for the analysis: valid height.

<b>Risk factor</b>	<b>Definition</b>	<b>Male prevalence</b>	<b>Female prevalence</b>	<b>Source:</b>	<b>Population:</b>
<b>7. Physical activity</b>	3 or more hours total leisure-time physical activity per week.	37.0%	23.0%	Physical statistic 2015. Report by the British Heart Foundation Centre on Population Approaches for Non-Communicable Disease Prevention, Nuffield Department of Population Health, University of Oxford.	UK population. Guidelines issued by the Chief Medical Officers (CMOs) of England, Scotland, Wales and Northern Ireland in 2011. Over a week, activity should add up to at least 150 minutes (2½ hours) of moderate intensity activity in bouts of 10 minutes or more.
<b>8. Red meat</b>	Eating 3 or more servings a week. (US meat one serving = 4 ounces or = 113.4 grams)	97.5%	91.0%	Parkin DM (2011) 5. Cancers attributable to dietary factors in the UK in 2010. II. Meat consumption. Br J Cancer 105 Suppl 2: S24-26.	Data on consumption of meat in the UK year 2000–2001 from the National Diet and Nutrition Survey (Food Standards Agency, 2002) as mean consumption, in grams of different types of meat per week, by age group and sex.
<b>9. Use of birth control pills</b>	5 or more years of use	N/A	57.0%	Farrow A, Hull MGR, Northstone K, Taylor H, Ford WCL, Golding J (2002) Prolonged use of oral contraception before a planned pregnancy is associated with a decreased risk of delayed conception. Human Reproduction 17(10): 2754-2761.	The Avon Longitudinal Study of Parents and Children (ALSPAC).
<b>10. Use of postmenopausal hormones</b>	5 or more years of use	N/A	8.0%	Benson VS, Kirichek O Fau - Beral V, Beral V Fau - Green J, Green J Menopausal hormone therapy and central nervous system tumor risk: large UK prospective study and meta-analysis. (1097-0215 (Electronic)).	UK population (General Practice Research Database (GPRD)).
<b>11. Aspirin use</b>	Use daily for 15 years or more.	22.0%	13.0%	Elwood P, Morgan G, White J, Dunstan F, Pickering J, Mitchell C, Fone D (2011) Aspirin taking in a south Wales county. The British Journal of Cardiology 18: 238-240.	Sample of adults residing in the south Wales county of Caerphilly, the study conducted a survey of a sample 9,551 adults resident in the county aged ≥18 years.
<b>12. Multivitamin (folate)</b>	folate intake reflected in regular multivitamin use (>15yrs vs no use)	5.0%	7.0%	Comparison of standardised dietary folate intake across ten countries participating in the European Prospective Investigation into Cancer and Nutrition. British Journal of Nutrition (2012), 108, 552–569.	UK population

<b>Risk factor</b>	<b>Definition</b>	<b>Male prevalence</b>	<b>Female prevalence</b>	<b>Source:</b>	<b>Population:</b>
<b>13. Inflammatory bowel disease</b>	Affected by the condition for 10 or more years.	1.0%	3.0%	Canavan C, Card T, West J (2014) The incidence of other gastroenterological disease following diagnosis of irritable bowel syndrome in the UK: a cohort study. PLoS One 9(9): e106478.	UK population (General Practice Research Database (GPRD)).
<b>14. Calcium supplement</b>	regular use of calcium supplement everyday	11.0%	11.0%	<a href="http://www.telegraph.co.uk/news/health/news/11900171/Calcium-supplements-dont-work-say-experts.html">http://www.telegraph.co.uk/news/health/news/11900171/Calcium-supplements-dont-work-say-experts.html</a>	UK population, up to 11 per cent of British adults are estimated take calcium supplements.
<b>16. Vitamin D supplement</b>	regular use of calcium supplement everyday	15.0%	15.0%	Spiro A and Buttriss J.L. (2014) Vitamin D:An overview of vitamin D status and intake in Europe. Nutrition Bulletin Vol 39;4.	Irish population

Table 2. Relative risks for selected breast cancer risk factors extracted from the original publication by Colditz et al(14) and from the UKBiobank data

<b>Risk factor</b>	<b>RR*</b>	<b>RR**</b>	<b>95% C.I.**</b>	<b>Score assigned in Colditz et al.</b>	<b>Score assigned in the UKBiobank study</b>
Family history (mother and sister)	3.0	3.0	2.0 -3.6	25	25
Family history (first-degree relative)	1.8	1.5	1.4-1.7	10	10
Height	1.3	1.3	1.3-1.5	5	5
Age of first period	0.8	0.9	0.9-1.0	-5	-5
Age at menopause	1.2	1.3	1.2-1.4	5	5
OC use	1.4	1.1	1.0-1.2	5	5
Estrogen replacement ≥5years	1.7	1.3	1.2-1.4	10	5
Estrogen replacement <5years	1.1	1.2	1.0-1.3	5	5
Physical activity	0.6	0.9	0.8-0.9	10	5
Alcohol	1.4	1.1	1.0-1.2	5	5
Obesity (postmenopausal)	1.3	1.1	1.1-1.2	5	5
Obesity (premenopausal)	0.8	0.8	0.8-1.0	-5	-5
Multivitamin supplement	0.5	0.9	0.9-1.0	10	0
No. of births	1.1	1.23	1.13-1.24	5	5
Benign breast disease (MD diagnosed)	1.5	1.4	1.1-1.8	10	5
birth weight	1.1	1.1	0.9-1.2	5	5

\*Relative risk from publication by Colditz et al, \*\* Relative risk from the UKBiobank study

Table 3. Risk score applied to level of relative risk

<b>Relative risk</b>	<b>Risk score</b>
0.9 to < 1.1	0
0.7 to < 0.9 or 1.1 to < 1.5	5
0.4 to < 0.7 or 1.5 to < 3.0	10
0.2 to < 0.4 or 3.0 to < 7.0	25
< 0.2 or $\geq$ 7.0	50

Table 4. Illustration of population average score and individual risk score for breast cancer

Risk factor	RR	Description	Score	UK prevalence women	Population average points for women	Individual factor profile	Individual score
Family history (mother and sister)	3.0	Two first degree relatives (mother and sister) affected with breast cancer before aged 65	25	0.0011	0.03	No	0
Family history (first-degree relative)	1.8	First degree relative who has a history of breast cancer before age 65 vs. none	10	0.09	0.90	No	0
Height	1.3	5 feet 7 inch or taller for women	5	0.07	0.35	Yes	5
Age of first period	0.8	Age of first period (15 vs.11)	-5	0.17	-0.85	Age 11	0
Age at menopause	1.2	Age at menopause (at the age of 55 or older)	5	0.07	0.34	Age 55	5
OC use	1.4	OC use (current use vs. none)	5	0.29	1.45	Yes	5
Oestrogen replacement	1.7	Oestrogen replacement $\geq$ 5yrs	10	0.08	0.80	No	0
Oestrogen replacement	1.1	Oestrogen replacement $<$ 5yrs	5	0.08	0.40	No	0
Physical activity	0.6	3 or more hours total leisure-time physical activity per week	10	0.77	7.70	No	10
Jewish heritage	1.2	Jewish heritage	5	0.005	0.03	No	0
Alcohol	1.4	More than 1 drink per day vs 0	5	0.41	2.05	Yes	5
Obesity (postmenopausal)	1.3	27 kg/m <sup>2</sup> or more	5	0.49	2.45	No	0
Obesity (premenopausal)	0.8	27 kg/m <sup>2</sup> or more	-5	0.36	-1.80	Yes	-5
Multivitamin supplement	0.7	lack of use of multivitamin or B complex	5	0.73	7.3	Use vitamin	0
No. of births	1.1	No. of births (0 or 1 child)	5	0.24	1.20	No child	0
Benign breast disease (MD diagnosed)	1.5	Benign breast disease (MD diagnosed)	10	0.13	1.30	No	0
Tamoxifen or raloxifene	0.5	tamoxifen or raloxifene for 5 years or more	10	0.30	3.0	No	10
Birth weight	1.1	birth weight $>$ 3.9kg or more	5	0.07	0.37	No	0
				Population average score	<b>27</b>	Individual risk score	<b>35</b>

Table 5. Conversion of individual index score to single numeric factor

<b>individual index score →</b>	<b>Level of risk →</b>	<b>factor</b>
<0	Very much below average risk	0.2
0, or < 0.5	Much below average risk	0.4
0.5 < 0.9	Below average risk	0.7
0.9 < 1.1	About average risk	1
1.1 < 2.0	Above average risk	1.5
2.0 < 5.0	Much above average risk	3
5.0 or more times the average score	Very much above average risk	5

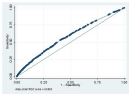


Table 6. 10-years estimated breast cancer risk in UK and US female population

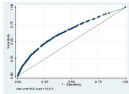
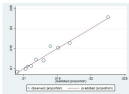
Age group	10-years estimated risk	
	UK females	US females*
0-9	0.000%	0.00%
10-19	0.001%	0.00%
20-29	0.049%	0.06%
30-39	0.442%	0.44%
40-49	1.445%	1.44%
50-59	2.594%	2.28%
60-69	3.336%	3.46%
70-79	2.759%	3.89%
80+	2.944%	3.02%

\* Estimates from US SEER data

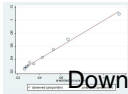
Figure1. Model performance of the breast cancer sub-component of the model scored using binary factor categorisation: AUC and calibration curves



1. Controls (no cancers)



2. Controls (no cancers/no other illnesses)



Download

# Cancer Prevention Research

## Development of a cancer risk prediction tool for use in the UK Primary Care and community settings.

Artitaya Lophatananon, Juliet A. Usher-Smith, Jackie Campbell, et al.

*Cancer Prev Res* Published OnlineFirst May 30, 2017.

<b>Updated version</b>	Access the most recent version of this article at: doi: <a href="https://doi.org/10.1158/1940-6207.CAPR-16-0288">10.1158/1940-6207.CAPR-16-0288</a>
<b>Author Manuscript</b>	Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, contact the AACR Publications Department at [permissions@aacr.org](mailto:permissions@aacr.org).