

## **Prospective Evaluation of a Breast Cancer Risk Model Integrating Classical Risk Factors and Polygenic Risk in 15 Cohorts from Six Countries**

Amber N Wilcox, MPH<sup>1,2\*</sup>, Parichoy Pal Choudhury, PhD<sup>1,3\*</sup>, Chi Gao, MS<sup>4,5</sup>, Anika Hüsing, PhD<sup>6</sup>, Mikael Eriksson, MSc<sup>7</sup>, Min Shi, PhD<sup>8</sup>, Christopher Scott, MS<sup>9</sup>, Brian D Carter, MPH<sup>10</sup>, Kara Martin, PhD<sup>11</sup>, Elaine Harkness, PhD<sup>12,13</sup>, Mark N Brook, PhD<sup>14</sup>, Thomas U Ahearn, PhD<sup>1</sup>, Nasim Mavaddat, PhD<sup>15</sup>, Antonis C Antoniou, PhD<sup>15</sup>, Jenny Chang-Claude, PhD<sup>6,16</sup>, Jacques Simard, PhD<sup>17</sup>, Michael E Jones, PhD<sup>14</sup>, Nick Orr, PhD<sup>18</sup>, Minouk J Schoemaker, PhD<sup>14</sup>, Anthony J Swerdlow, DSc<sup>14,19</sup>, Sarah Sampson, BS<sup>20</sup>, William G Newman, PhD<sup>21,22</sup>, Elke M van Veen, MSc<sup>21,22</sup>, D. Gareth R Evans, MD<sup>21,22</sup>, Robert J MacInnis, PhD<sup>11,23</sup>, Graham G Giles, PhD<sup>11,23,24</sup>, Melissa Southey, PhD<sup>25,26,11</sup>, Roger L Milne, PhD<sup>11,23,25</sup>, Susan M Gapstur, PhD<sup>10</sup>, Mia M Gaudet, PhD<sup>10</sup>, Stacey J Winham, PhD<sup>27</sup>, Kathy Brandt, MD<sup>28</sup>, Aaron Norman, PhD<sup>9</sup>, Celine M Vachon, PhD<sup>29</sup>, Dale P Sandler, PhD<sup>30</sup>, Clarice R Weinberg, PhD<sup>8</sup>, Kamila Czene, PhD<sup>7</sup>, Marike Gabrielson, PhD<sup>7</sup>, Per Hall, PhD<sup>7,31</sup>, Carla H van Gils, PhD<sup>32</sup>, Kay-Tee Khaw, MBBChir<sup>33</sup>, Myrto Barrdahl, PhD<sup>6</sup>, Rudolf Kaaks, PhD<sup>6</sup>, Paul M Ridker, MD, MPH<sup>34,35</sup>, Julie E Buring, ScD<sup>34,35</sup>, Dan I Chasman, PhD<sup>34,35</sup>, Douglas F Easton, PhD<sup>36,15</sup>, Marjanka K Schmidt, PhD<sup>37,38</sup>, Peter Kraft, PhD<sup>4,5</sup>, Montserrat Garcia-Closas, MD, DrPH<sup>1\*\*</sup>, Nilanjan Chatterjee, PhD<sup>39,40\*\*</sup>

### **Affiliations:**

<sup>1</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Dr, Rockville, MD, 20850, US, <sup>2</sup>Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, 135 Dauer Drive, Chapel Hill, NC, 27599, US, <sup>3</sup>Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 615 N Wolfe Street, Baltimore, MD, 21205, US, <sup>4</sup>Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA, 02115, US, <sup>5</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA, 02115, US, <sup>6</sup>Division of Cancer Epidemiology,

German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg, 69120, Germany,<sup>7</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Karolinska Univ Hospital, Stockholm, 171 65, Sweden,<sup>8</sup>Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, NIH, 111 T.W. Alexander Drive, Research Triangle Park, NC, 27709, US,<sup>9</sup>Department of Health Sciences Research, Mayo Clinic, 200 First Street SW, Rochester, MN, 55905, US,<sup>10</sup>Behavioral and Epidemiology Research Group, American Cancer Society, 250 Williams Street NW, Atlanta, GA, 30303, US,<sup>11</sup>Cancer Epidemiology Division, Cancer Council Victoria, 615 St Kilda Road, Melbourne, Victoria, 3004, Australia,<sup>12</sup>Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Oxford Road, Manchester, M13 9PT, UK,<sup>13</sup>Nightingale Breast Screening Centre, Wythenshawe Hospital, Manchester University NHS Foundation Trust, Southmoor Road, Manchester, M23 9LT, UK,<sup>14</sup>Division of Genetics and Epidemiology, The Institute of Cancer Research, 15 Cotswold Road, London, SM2 5NG, UK,<sup>15</sup>Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, 2 Worts' Causeway, Cambridge, CB1 8RN, UK,<sup>16</sup>Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Martinistraße 52, Hamburg, 20246, Germany,<sup>17</sup>Centre Hospitalier Universitaire de Québec–Université Laval Research Center, 2705 Laurier Boulevard, Québec City, QC, G1V 4G2, Canada,<sup>18</sup>Centre for Cancer Research and Cell Biology, Queen's University Belfast, University Road, Belfast, Ireland, BT7 1NN, UK,<sup>19</sup>Division of Breast Cancer Research, The Institute of Cancer Research, 123 Old Brompton Road, London, SW7 3RP, UK,<sup>20</sup>Prevention Breast Cancer Unit and Nightingale Breast Screening Centre, Manchester University NHS Foundation Trust (South), Southmoor Road, Manchester, M23 9LT, UK,<sup>21</sup>Division of Evolution and Genomic Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Oxford Road, Manchester, M13 9WL, UK,<sup>22</sup>Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester NIHR Biomedical Research Centre, Manchester University Hospitals NHS, Foundation Trust, Manchester Academic Health Science Centre, Oxford Road, Manchester, M13 9WL, UK,<sup>23</sup>Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Level 1, 723 Swanston Street, Melbourne, Victoria, 3010, Australia,<sup>24</sup>Department of Epidemiology and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, Victoria, 3004, Australia,<sup>25</sup>Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, 246 Clayton Road, Clayton, Victoria, 3168, Australia,<sup>26</sup>Department of Clinical Pathology, The University of Melbourne, Cnr Grattan Street and Royal Parade, Melbourne, Victoria, 3010, Australia,<sup>27</sup>Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, 200 First Street SW, Harwick 6, Rochester, MN, 55905, US,<sup>28</sup>Department of Radiology, Mayo Clinic, 200 First St SW Mayo Building East 2, Rochester, MN, 55905, US,<sup>29</sup>Department of Health Sciences Research, Division of Epidemiology, Mayo Clinic, 200 First Street SW, Harwick 6, Rochester, MN, 55905, US,<sup>30</sup>Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, 111 T.W. Alexander Drive, Research Triangle Park, NC, 27709, US,<sup>31</sup>Department of Oncology, Södersjukhuset, Sjukhusbacken 10, Stockholm, 118 83, Sweden,<sup>32</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Heidelberglaan 100, Utrecht, 3584 CX, The Netherlands,<sup>33</sup>Cancer Epidemiology Unit, University of Cambridge, Robinson Way, Cambridge, CB20QQ, UK,<sup>34</sup>Division of Preventive

Medicine, Brigham and Women's Hospital, 900 Commonwealth Avenue, Boston, MA, 02115, US, <sup>35</sup>Harvard Medical School, 25 Shattuck Street, Boston, MA, 02115, US, <sup>36</sup>Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, 2 Worts' Causeway, Cambridge, CB1 8RN, UK, <sup>37</sup>Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Plesmanlaan 121, Amsterdam, 1066 CX, The Netherlands, <sup>38</sup>Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Plesmanlaan 121, Amsterdam, 1066 CX, The Netherlands, <sup>39</sup>Bloomberg School of Public Health, Johns Hopkins University, 615 N Wolfe Street, Baltimore, MD, 21205, US, <sup>40</sup>Department of Oncology, School of Medicine, Johns Hopkins University, 733 N Broadway, Baltimore, MD, 21205, US.

The authors have no conflicts of interest to disclose.

\*Co-first authors

\*\*Co-last authors

## **SUPPORT**

BCAC is funded by Cancer Research UK [C1287/A16563, C1287/A10118], the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and by the European Community's Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS). The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report. Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer UK Grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme

under grant agreement number 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defense (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, and Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The DRIVE Consortium was funded by U19 CA148065. The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Éducation Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk,

MR/M012190/1 to EPIC-Oxford) (United Kingdom). The KARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer. MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 209057, 396414, and 1074383, and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database. The MMHS study was supported by NIH grants CA97396, CA128931, CA116201, CA140286 and CA177150. The NHS was supported by NIH grants P01 CA87969, UM1 CA186107, and U19 CA148065. The NHS2 was supported by NIH grants UM1 CA176726 and U19 CA148065. Genotyping for PLCO was supported by the Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. The PLCO is supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, National Institutes of Health. PROCAS is funded from NIHR grant PGfAR 0707-10031. The Sister Study (SISTER) is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES049033). The UKBGS is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. ACA is supported by Cancer Research-UK (C12292/A20861). Data analysis and iCARE package development was supported by the Patient-Centered Outcomes Research Institute (PCORI) Award (ME1602-34530). ANW is supported by the University of North Carolina LCCC Cancer Control Education Program predoctoral training grant from the National Cancer Institute (T32-CA057726). MGC and PPC are supported by the Intramural Research

Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, USA.

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. The COGS study would not have been possible without the contributions of the following: Paul Pharoah, Kyriaki Michailidou, Manjeet K. Bolla, Qin Wang (BCAC), Andrew Berchuck (OCAC), Rosalind A. Eeles, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRACTICAL), Georgia Chenevix-Trench, Lesley McGuffog, Fergus Couch and Ken Offit (CIMBA), Joe Dennis, Alison M. Dunning, Andrew Lee, and Ed Dicks, Craig Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory, Javier Benitez, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, Stig E. Bojesen, Sune F. Nielsen, Borge G. Nordestgaard, and the staff of the Copenhagen DNA laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility. Investigators from the CPS-II cohort thank the participants and Study Management Group for their invaluable contributions to this research. They also acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, as well as cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program. We thank the participants and the investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). KARMA thanks the Swedish Medical Research Counsel. We thank the coordinators, the research staff and especially the MMHS participants for their continued collaboration on research studies in breast

cancer. We would like to thank the participants and staff of the NHS and NHS2 for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data. PROCAS thanks NIHR for funding. DGE and EFH are supported by the NIHR Manchester Biomedical Research Centre (IS-BRC-1215-20007). UKBGS thanks Breast Cancer Now and the Institute of Cancer Research for support and funding of the Breakthrough Generations Study, and the study participants, study staff, and the doctors, nurses and other health care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR and Manchester NIHR Biomedical Research Centres. The WGHS is supported by the National Heart, Lung, and Blood Institute (HL043851 and HL080467) and the National Cancer Institute (CA047988 and UM1CA182913), with funding for genotyping provided by Amgen.

### **CORRESPONDING AUTHOR**

Dr. Montserrat Garcia-Closas

Deputy Director, Division of Cancer Epidemiology & Genetics

National Cancer Institute | Shady Grove Campus

DCEG Office of the Director, Room 7E342

9609 Medical Center Drive Rockville, MD 20850

T +1 240-276 7648 | E-mail [montserrat.garcia-closas@nih.gov](mailto:montserrat.garcia-closas@nih.gov)

**RUNNING HEAD:** Prospective evaluation of breast cancer risk model in 6 countries

**PRIOR PRESENTATION:** American Association for Cancer Research Annual Meeting,

March

2019

## **ABSTRACT**

### **PURPOSE**

Risk-stratified breast cancer prevention requires accurate identification of women at sufficiently different levels of risk. We conducted a comprehensive evaluation of a model integrating classical risk factors and a recently developed 313-variant polygenic risk score (PRS) to predict breast cancer risk.

### **METHODS**

Fifteen prospective cohorts from six countries with 237,632 women (7,529 incident breast cancer patients) of European ancestry aged 19-75 years at baseline were included. Calibration of five-year risk was assessed by comparing predicted and observed proportions of cases overall and within risk categories. Risk stratification for women of European ancestry aged 50-70 years in those countries was evaluated by the proportion of women and future breast cancer cases crossing clinically-relevant risk thresholds.

### **RESULTS**

The model integrating classical risk factors and PRS accurately predicted five-year risk. For women younger than 50 years, median (range) expected-to-observed ratio across the cohorts was 0.94 (0.72 to 1.01) overall and 0.9 (0.7 to 1.4) at the highest risk decile. For women 50 years or older, these ratios were 1.04 (0.73 to 1.31) and 1.2 (0.7 to 1.6), respectively. The proportion of women in the general population identified above the 3% five-year risk threshold (used for recommending risk-reducing medications in the US) ranged from 7.0% in Germany (~841,000 of 12 million) to 17.7% in the US (~5.3 of 30 million). At this threshold, 14.7% of US women were re-classified by the addition of PRS to classical risk factors, identifying 12.2% additional future breast cancer cases.



## CONCLUSION

Evaluation across multiple prospective cohorts demonstrates that integrating a 313-SNP PRS into a risk model substantially improves its ability to stratify women of European ancestry for applying current breast cancer prevention guidelines.

## INTRODUCTION

Clinical guidelines for breast cancer prevention and early detection use risk thresholds to identify women eligible for interventions, e.g., genetic counseling, tailored screening, risk-reducing drugs or surgery.<sup>1-3</sup> Risk assessment under these guidelines can be based on age, family history of cancer, mutation status for breast cancer genes such as *BRCA1* and *BRCA2*, history of benign breast disease, and other risk factors. Several risk models are available to estimate absolute risk, each including different sets of risk factors and aimed at different clinical scenarios.<sup>4,5</sup> For instance, family-based models using pedigree-level information on family history of cancer and mutation testing are typically used for genetic counseling of women in cancer genetics clinics, whereas models with more limited family history information and personal risk factors are often used to identify women for referral to genetic or oncology clinics. Many of these models are used clinically with limited evidence on their comparative performance across populations.<sup>6-8</sup> Additionally, there is a need to improve the risk discrimination of current models, particularly those assessing risk in the general population.

Polygenic risk scores (PRS) based on common single nucleotide polymorphisms (SNPs), identified by genome-wide association studies (GWAS)<sup>9</sup> can provide improved risk stratification of breast cancer, i.e., allow accurate identification of higher proportion of populations at the extremes, compared to other risk factors for women with or without family history.<sup>8,10-14</sup> Risk

assessment using PRS is already commercially available and marketed to clinicians.<sup>15,16</sup> However, while PRSs alone have been demonstrated to provide accurate relative risks for breast cancer in women of European ancestry,<sup>17</sup> evidence on the accuracy of predicted risk from models integrating PRS and classical risk factors in prospective cohort studies is very limited.<sup>7,8,11,16,18</sup> Thus, there is an urgent need for this type of assessment in large cohorts across multiple populations.

Individualized Coherent Absolute Risk Estimator (iCARE) is a new flexible tool<sup>19</sup> to develop and validate risk prediction models in the general population. The iCARE models for breast cancer based on classical risk factors showed similar or better prediction accuracy and discrimination compared to two commonly used models in research and clinical settings, BCRAT and IBIS, in two large prospective cohorts.<sup>8</sup> The current study aims to conduct an extensive independent validation of iCARE-based models that integrate classical risk factors and a 313-SNP PRS<sup>17</sup> in 15 prospective cohorts across six countries. We also examined risk stratification of these models by estimating five-year risk projections in those countries. Finally, we provided projections for improved levels of risk stratification attainable by PRSs expected to result from ongoing efforts to double the size of current breast cancer GWAS.<sup>20</sup>

## **METHODS**

### **Study Populations**

Eligible person-time within the 15 cohorts began at the time of DNA collection for women of self-reported European ancestry aged 18-75 years with no personal history of breast cancer, who consented for genetic studies and completed a risk factor questionnaire. For studies with multiple questionnaires, we used those administered closest to DNA collection (Table S1). There was a

wide range of variation in the year and average age of DNA collection across the populations (Table S1). The distribution of risk factors varied substantially across the cohorts (Tables S2-S3), across the underlying country-specific reference populations, and between cohorts and the reference population within the same country (Tables S4-S5).

### **Risk Models**

We used iCARE to build models for five-year absolute risk of developing breast cancer<sup>19</sup> integrating classical risk factors and PRS (Figure 1). Age was used as the timescale in disease incidence modeling. The conditional age-specific incidence rates given the risk factors were assumed to follow a Cox proportional hazards model.<sup>21</sup> Classical risk factors in the iCARE-Lit model were specified as previously described.<sup>8</sup> Briefly, this model includes ages at menarche and first birth, parity, height, alcohol intake, breast cancer family history (i.e., presence or absence of disease in at least one first-degree relative), history of benign breast disease, oral contraceptive use and BMI for all women. For women 50 years or older it additionally includes age at menopause, hormone replacement therapy (HRT) use, and current HRT type. The relative risks for these factors were obtained from a literature review separately for women younger and older than 50 years (Supplements).<sup>8</sup> The 313-SNP PRS, recently described in Mavaddat et al.,<sup>17</sup> was integrated into the iCARE-Lit model assuming a multiplicative joint association with disease risk, except for family history where we accounted for attenuation of the relative risk due to its correlation with PRS.

To estimate absolute risk of breast cancer, we incorporated country-specific incidence and competing mortality rates from population-based registries and the population distribution of risk factors, using an individual-level reference dataset representative of each population (Figure S1,

Tables S4-S5). We simulated the PRS in the reference dataset assuming independence with classical risk factors, conditional on family history.<sup>19</sup>

### **Cohort-Specific Validation of Risk Models**

Women developing incident primary breast cancer, either in-situ or invasive, during the follow-up period were considered cases (Table S1). Follow-up for all subjects was assumed to begin one year following study entry to reduce the impact of prevalent cases on validation analyses and was defined up to the last contact or record linkage to cancer registry or five years, whichever came first. Predicted risk estimates account for competing mortality due to other causes during the follow-up period (Figure 1). The model validation methods are discussed in detail in our previous work<sup>8</sup> and the Supplements.

### **Meta-Analysis of Relative Risk Calibration**

We evaluated relative risk calibration of each model through a meta-analysis across cohorts using the deciles of the relative risk score in UK Biobank as the standard set of cutoffs to delineate the risk categories. We first computed the relative risks (observed and expected) for each decile with respect to the fifth decile and combined the estimates using standard inverse variance weighted meta-analysis<sup>22</sup> on the log-relative risk scale. We also computed a meta-analysis estimator and a Wald-statistic based 95% confidence interval (CI) of the AUC based on all risk factors except age.

### **Projection of Risk to the General Populations**

Projected five-year absolute risks for women of European ancestry aged 50-70 years in each country were calculated using the population distribution of ages and risk factors in the corresponding reference dataset, population-specific breast cancer incidence and competing mortality rates (Figure 1, Supplements). Proportions of women and incident cases expected to

arise within five years were calculated at both extremes of the risk distribution based on two high-risk thresholds and two low-risk thresholds. Low-risk cutoffs were 0.6% and 1.13%, corresponding to average five-year risk for US women aged 40 and 50 years, respectively. High-risk cutoffs were 3%, corresponding to US Preventive Services Task Force recommendation for risk-lowering drugs<sup>2</sup>, and 6%, corresponding to a cutoff for very high risk used in the WISDOM trial.<sup>23</sup> We also evaluated expected improvements in risk stratification by incorporating an improved PRS to the iCARE models (Supplements).

### **Reclassification Analysis**

We evaluated potential improvements in risk stratification achieved by adding genetic variants to the model with classical risk factors in the populations of the UK and the US in terms of proportions of women and future cases reclassified at the extremes of the risk distribution based on two low-risk (0.6% and 1.13%) and two high-risk cutoffs (3% and 6%). Women who were at 8% or higher risk, corresponding to the National Institute for Health and Care Excellence (NICE) guidelines for ten-year risk of a 40-year-old woman in the UK,<sup>3</sup> were above 3% five-year risk of breast cancer.

### **RESULTS**

Meta-analyses of the iCARE-Lit model relative risk scores showed that observed and expected scores based on the 313-SNP PRS were not significantly different overall, with excellent calibration across risk deciles (Figure S2). While relative risk scores based on classical risk factors appeared well calibrated for most cohorts (Figures S3-S5), the meta-analysis showed 45% and 18% overestimation of risk for the highest decile for women younger and older than 50 years, respectively, and 16% underestimation of risk for women in the lowest decile, only in the older group (Figure S2). The integrated model also appeared well calibrated across cohorts

(Figures S3-S5), and meta-analyses showed good calibration across most risk categories, but a 40% and 27% overestimation in the highest risk decile for the younger and older age groups, respectively (Figure 2, Figure S2). The analyses for women older than 50 years using a similar model with relative risk estimates from a multivariable analysis of prospective cohorts (iCARE-BPC3<sup>24</sup>) provided somewhat better calibration of models at the high-risk decile (Figure S2, Figure S6).

For both age groups, incorporating the 313-SNP PRS to the classical risk factors substantially improved overall risk discrimination (Figure S2). Age-adjusted AUC (95% CI) was 55.9 (53.8 to 58.0) and 64.0 (62.0 to 66.0) for classical risk factors and integrated models, respectively, in younger women; and 57.3 (56.5 to 58.1) and 64.0 (63.2 to 64.7), respectively, for older women (Figure 2). However, the differences between the AUCs based on the 313-SNP PRS alone and integrated models were smaller, with overlap in 95% CIs (Figure S2). All the models showed improvements in AUC after incorporating the variation due to age at enrollment (Figure 3, Figures S11-S13).

While calibration of absolute risk for the integrated models varied substantially across cohorts (Figures 4-5, S7A-S10B), the models did not systematically over- or under-predict risk. The median (range) expected-to-observed (E/O) ratio across the studies were 0.94 (0.72 to 1.01) for women younger than 50 years and 1.04 (0.73 to 1.31) for women older than 50 years. For women at the highest risk decile these were 0.9 (0.7 to 1.4) and 1.2 (0.7 to 1.6) for the two age groups, respectively.

Risk projections among women of European ancestry aged 50-70 years showed that addition of PRS to the classical risk factors produced substantial improvements in risk stratification (Figure 6, Tables S7A-S7F), resulting in identifying a large fraction of the population at high or low risk

(five-year risk  $>3\%$  or  $\leq 1.13\%$ , respectively). These projections vary across populations due to differences in the distribution of age and classical risk factors. The proportion of women identified at low risk ranged from 27.0% in the US (~8.1 of 30 million) to 46.0% in Germany (~5.5 of 12 million). At the high-risk threshold, the proportion of women identified ranged from 7.0% in Germany (~841,000 of 12 million) to 17.7% in the US (~5.3 of 30 million). As a sensitivity analysis, we adjusted for the miscalibration of risk in the high-risk decile (Figure 2) by attenuating the risk projections by the amount of overestimation in that category. After this adjustment, the proportion of women identified at the high-risk threshold ranged from 5.6% in Germany (~677,000 of 12 million) to 16.7% in the US (~5 of 30 million) (Table S8). Proportions of future cases identified at high risk also remained substantial after adjusting for the relative risk miscalibration.

Our projections indicate that incorporating an improved PRS substantially increases the number of future cases expected to be identified in the high-risk group, with only a marginal predicted increase in total women in that group. In the low-risk group, an improved model would identify a substantially higher fraction of total women, with only a minimal number of additional cases expected (Tables S7A-S7F).

Addition of the 313-SNP PRS to the classical risk factors (Table 1, Table S6) resulted in a substantial reclassification of women leading to a net increase in the proportion of cases crossing risk thresholds under current clinical guidelines for breast cancer prevention. For example, for a five-year risk threshold of 3% recommended for risk-reducing medications in the US, 14.7% of US women were re-classified by the integrated model, leading to identification of an additional 12.3% of the future breast cancer cases. For the same threshold, the integrated model reclassified 15.3% of UK women, resulting in identifying an additional 15.7% of the future cases.

## DISCUSSION

The integrated iCARE-Lit model with 313-SNP PRS and classical risk factors showed substantial improvements in breast cancer risk prediction across 15 prospective cohorts of women of European ancestry in six countries. The integrated models generally showed good calibration and provided wider stratification of population risk, identifying more women crossing clinically actionable risk thresholds used by current guidelines for breast cancer prevention and early detection.

The integrated iCARE-Lit model provided well-calibrated relative risk scores across validation cohorts (Figures S3-S5); however, meta-analyses showed some over-prediction in the highest risk decile that was driven by the classical risk factor component. The calibration of absolute risks, however, varied more widely across cohorts, even within countries, though we did not see evidence of systematic under- or over-prediction across studies. This suggests that differences across cohorts are likely due to random variation or differences between study populations (e.g., wide range of study time periods (1989-2013) and differences in risk factor distributions or disease rates), rather than a reflection of intrinsic model properties. This highlights the importance of absolute risk validation across multiple study populations, particularly using cohorts similar to the target populations, both in chronologic years of study and underlying risk. Further studies in countries represented here by only one cohort, or not included in this report, are needed to evaluate country-specific differences in model performance.

We recently showed that five-year risk predictions by the iCARE-Lit model based on classical risk factors were at least as accurate as two established models used in clinical practice: BCRAT (“Gail”) and IBIS (“Tyrrer-Cuzick”).<sup>8</sup> Previous evaluations (e.g., Terry et al. 2019<sup>6</sup>) of BCRAT



and IBIS also showed overestimation of risk in the highest risk categories.<sup>25-27</sup> Our analyses suggested that building models from multivariable analysis of classical risk factors in prospective cohorts<sup>24</sup>, rather than from the literature, could improve calibration at the high-risk decile. Thus, calibration and discrimination of models can potentially improve through efforts of building multivariate relative risk models in prospective cohorts of women across age groups, and with more comprehensive information on questionnaire-based risk factors. Addition of risk biomarkers such as mammographic breast density,<sup>11,12,27,28</sup> circulating hormone levels,<sup>28-30</sup> or novel risk factors as they are identified in the future can result in further improvements.

Discriminatory accuracy of risk models may be substantially different in research cohorts than in target populations due to differences in underlying risk factor distributions. For instance, our projections show a higher discriminatory accuracy in the US population (AUC = 66.5, Table S7F, Figure S14) compared to the US-based cohorts (AUC range: 63.1-65.8) (Figures S11-S12).

Our projections also show that an improved PRS, achievable through larger GWAS, could lead to better risk stratification, with a model integrating risk factors and an improved PRS achieving an AUC~0.71 (Tables S7A-S7F). This will improve our ability to identify women eligible for risk-reducing interventions, or supplemental screening by magnetic resonance imaging or other imaging modalities. However, since the discriminatory performance of models will remain moderate, most breast cancers will still occur among women not identified at elevated risk. Thus, broader public health efforts targeting the whole population will continue to be required for reducing the population burden of breast cancer in a major way.<sup>31</sup> Risk-stratified screening strategies at the population level tailored to women's individual risks based on integrated models may improve the effectiveness of population-based screening, relative to the current age-stratified programs,<sup>32</sup> and is currently being evaluated in screening trials<sup>33</sup>.

We used country-specific breast cancer incidence rates and a reference dataset for each country, built from population-based surveys, to translate relative risks to absolute risk estimates over a specified time period.<sup>19</sup> This critical step in building absolute risk models minimizes miscalibration of absolute risk when models developed for one target population are applied to other populations or countries. Although most model algorithms allow for changes in default incidence rates, the underlying risk factor distributions are often implicit and cannot be easily changed. While this flexibility is a strength of the iCARE modeling approach, it requires the availability of population-based survey data with information on all risk factors included in the models for the relevant time periods. We were unable to identify a single data source for the distribution of all the risk factors in each target population, requiring us to simulate some risk factors and make various modeling assumptions (e.g., independence of certain risk factors). This could have affected model performance across study populations. Finally, the reference datasets enable iCARE models to provide absolute risk estimates for individuals based on a subset of risk factors in the model.<sup>19</sup> This feature adds flexibility to use the models in different settings using information on a subset of risk factors.

Our risk models are aimed at the general population and do not adequately capture risk for women with strong family histories or carrying high-risk mutations. This requires integration with family-based models, e.g., our recent extension of the BOADICEA model<sup>13</sup> to include the iCARE-Lit risk factor component. However, this fully extended model has not yet been prospectively validated. Although iCARE can be used for risk predictions over any time period, the current study only evaluated five-year risk prediction, and further work is needed to evaluate longer-term predictions used by some clinical guidelines.

The iCARE-Lit model includes parameters for atypical hyperplasia, lobular carcinoma in situ and other benign breast diseases. However, this information was not available from the participating cohorts, thus further validation is required for risk prediction for women with these conditions. The iCARE models predict risk of overall breast cancer (i.e., invasive and *in situ*), rather than specific subtypes. Because risk factor associations and the effectiveness of preventive and screening strategies vary by tumor subtypes (e.g., estrogen receptor positive and negative tumors),<sup>34-39</sup> future work on subtype-specific risk predictions could result in more precise identification of women who would benefit most from specific interventions. Finally, the current models were derived and evaluated in studies of women of European ancestry and additional studies are urgently needed to develop and validate models for other populations, for whom alternative models have only been evaluated in relatively small studies.<sup>40-42</sup>

In summary, we present extensive validation results of a breast cancer risk prediction model integrating a newly developed PRS and classical risk factors. We show that it can provide substantial improvement in risk assessment for application of current clinical guidelines, or future risk-stratified prevention and screening strategies.<sup>43,44</sup>

## **AUTHOR CONTRIBUTIONS**

MGC and NC conceived the study. PPC developed and implemented the method and code for the statistical analyses. ANW coordinated analyses with the lead analysts from each cohort to obtain results. CG, AH, ME, MS, CS, BDC, KM, and EH analyzed data to generate preliminary results from the participating cohorts. ANW analyzed preliminary results from each study to generate the final version of the results presented in the main manuscript and supplementary materials. ANW, PPC, MGC and NC interpreted the findings and led the writing of the

manuscript. All authors contributed to the writing and reviewing of the final manuscript and its  
earlier draft.

## REFERENCES

1. Visvanathan K, Fabian CJ, Bantug E, et al: Use of Endocrine Therapy for Breast Cancer Risk Reduction: ASCO Clinical Practice Guideline Update. *Journal of Clinical Oncology* 0:JCO.19.01472
2. Owens DK, Davidson KW, Krist AH, et al: Medication Use to Reduce Risk of Breast Cancer: US Preventive Services Task Force Recommendation Statement. *Jama* 322:857-867, 2019
3. Familial breast cancer: classification, care and managing breast cancer and related risks in people with a family history of breast cancer. The National Institute for Health and Care Excellence (NICE). [nice.org.uk/guidance/cg164](http://nice.org.uk/guidance/cg164), 2013
4. Cintolo-Gonzalez JA, Braun D, Blackford AL, et al: Breast cancer risk models: a comprehensive overview of existing models, validation, and clinical applications. *Breast Cancer Res Treat* 164:263-284, 2017
5. Louro J, Posso M, Hilton Boon M, et al: A systematic review and quality assessment of individualised breast cancer risk prediction models. *Br J Cancer*, 2019
6. Terry MB, Liao Y, Whittemore AS, et al: 10-year performance of four models of breast cancer risk: a validation study. *Lancet Oncol* 20:504-517, 2019
7. Garcia-Closas M, Chatterjee N: Assessment of breast cancer risk: which tools to use? *Lancet Oncol* 20:463-464, 2019
8. Pal Choudhury P, Wilcox AN, Brook MN, et al: Comparative validation of breast cancer risk prediction models and projections for future risk stratification. *J Natl Cancer Inst*, 2019. In press
9. Michailidou K, Lindstrom S, Dennis J, et al: Association analysis identifies 65 new breast cancer risk loci. *Nature* 551:92-94, 2017
10. Garcia-Closas M, Gunsoy NB, Chatterjee N: Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. *J Natl Cancer Inst* 106, 2014
11. Vachon CM, Pankratz VS, Scott CG, et al: The contributions of breast density and common genetic variation to breast cancer risk. *J Natl Cancer Inst* 107, 2015
12. van Veen EM, Brentnall AR, Byers H, et al: Use of Single-Nucleotide Polymorphisms and Mammographic Density Plus Classic Risk Factors for Breast Cancer Risk Prediction. *JAMA Oncol* 4:476-482, 2018
13. Lee A, Mavaddat N, Wilcox AN, et al: BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med*, 2019
14. Shieh Y, Hu D, Ma L, et al: Breast cancer risk prediction using a clinical risk model and polygenic risk score. *Breast Cancer Res Treat* 159:513-25, 2016
15. Torkamani A, Wineinger NE, Topol EJ: The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics* 19:581-590, 2018
16. Sugrue LP, Desikan RS: What Are Polygenic Scores and Why Are They Important? *Jama* 321:1820-1821, 2019
17. Mavaddat N, Michailidou K, Dennis J, et al: Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* 104:21-34, 2019
18. Evans DGR, Harkness EF, Brentnall AR, et al: Breast cancer pathology and stage are better predicted by risk stratification models that include mammographic density and common genetic variants. *Breast Cancer Res Treat* 176:141-148, 2019

19. Pal Choudhury P, Maas P, Wilcox A, et al: iCARE: An R Package to Build, Validate and Apply Absolute Risk Models. *bioRxiv*, 2018. <https://doi.org/10.1101/079954>
20. Confluence Project, NCI Division of Cancer Epidemiology and Genetics
21. Cox DR: Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)* 34:187-220, 1972
22. Michael B, Larry V, Julian H, et al: Introduction to meta-analysis. UK, A John Wiley and Sons Ltd, 2009
23. Shieh Y, Eklund M, Madlensky L, et al: Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. *J Natl Cancer Inst* 109, 2017
24. Maas P, Barrdahl M, Joshi AD, et al: Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA Oncol* 2:1295-1302, 2016
25. Spiegelman D, Colditz GA, Hunter D, et al: Validation of the Gail et al. model for predicting individual breast cancer risk. *J Natl Cancer Inst* 86:600-7, 1994
26. Pastor-Barriuso R, Ascunce N, Ederra M, et al: Recalibration of the Gail model for predicting invasive breast cancer risk in Spanish women: a population-based cohort study. *Breast cancer research and treatment* 138:249-259, 2013
27. Brentnall AR, Cuzick J, Buist DSM, et al: Long-term Accuracy of Breast Cancer Risk Assessment Combining Classic Risk Factors and Breast Density. *JAMA Oncol* 4:e180174, 2018
28. Zhang X, Rice M, Tworoger SS, et al: Addition of a polygenic risk score, mammographic density, and endogenous hormones to existing breast cancer risk prediction models: A nested case-control study. *PLoS Med* 15:e1002644, 2018
29. Husing A, Fortner RT, Kuhn T, et al: Added Value of Serum Hormone Measurements in Risk Prediction Models for Breast Cancer for Women Not Using Exogenous Hormones: Results from the EPIC Cohort. *Clin Cancer Res* 23:4181-4189, 2017
30. Clendenen TV, Ge W, Koenig KL, et al: Breast cancer risk prediction in women aged 35–50 years: impact of including sex hormone concentrations in the Gail model. *Breast Cancer Research* 21:42, 2019
31. Gail MH, Pfeiffer RM: Breast Cancer Risk Model Requirements for Counseling, Prevention, and Screening. *J Natl Cancer Inst*, 2018
32. Pashayan N, Morris S, Gilbert FJ, et al: Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. *JAMA Oncol*, 2018
33. Antoniou A, Anton-Culver H, Borowsky A, et al: A response to "Personalised medicine and population health: breast and ovarian cancer". *Hum Genet* 138:287-289, 2019
34. Yang XR, Chang-Claude J, Goode EL, et al: Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. *J Natl Cancer Inst* 103:250-63, 2011
35. Holm J, Eriksson L, Ploner A, et al: Assessment of Breast Cancer Risk Factors Reveals Subtype Heterogeneity. *Cancer Res* 77:3708-3717, 2017
36. Brouckaert O, Rudolph A, Laenen A, et al: Reproductive profiles and risk of breast cancer subtypes: a multi-center case-only study. *Breast Cancer Research* 19:119, 2017
37. Bareche Y, Venet D, Ignatiadis M, et al: Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. *Ann Oncol* 29:895-902, 2018
38. Yeo SK, Guan J-L: Breast Cancer: Multiple Subtypes within a Tumor? *Trends in cancer* 3:753-760, 2017

39. Li K, Anderson G, Viallon V, et al: Risk prediction for estrogen receptor-specific breast cancers in two large prospective cohorts. *Breast Cancer Research*, 2018
40. Matsuno RK, Costantino JP, Ziegler RG, et al: Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. *J Natl Cancer Inst* 103:951-61, 2011
41. Banegas MP, John EM, Slattery ML, et al: Projecting Individualized Absolute Invasive Breast Cancer Risk in US Hispanic Women. *J Natl Cancer Inst* 109, 2017
42. McCarthy AM, Armstrong K, Handorf E, et al: Incremental impact of breast cancer SNP panel on risk classification in a screening population of white and African American women. *Breast Cancer Res Treat* 138:889-98, 2013
43. Personalised prevention in breast cancer: the policy landscape (2017) ISBN 978-1-907198-29-8,
44. Personalising prevention for breast cancer: workshop report (2019) ISBN 978-1-907198-34-2.

## FIGURE LEGENDS

**Figure 1.** Conceptual diagram of model building and validation of Individualized Coherent Absolute Risk Estimator (iCARE) for breast cancer. iCARE-BPC3 = Individualized Coherent Absolute Risk Estimation model based on Breast and Prostate Cancer Cohort Consortium, iCARE-Lit = iCARE model based on literature review.

**Figure 2.** Relative risk calibration of integrated breast cancer risk models (with classical risk factors and PRS) based on meta-analysis across validation studies. Classical risk factors include age at menarche, age at first live birth, parity, oral contraceptive use, age at menopause, hormone replacement therapy use, type of hormone replacement therapy, alcohol intake, height, BMI, breast cancer family history (i.e., presence or absence of breast cancer in at least one first degree relative), and benign breast disease. Meta-analysis is based on a reduced set of risk factors that were available in the majority of the validation cohorts. History of benign breast disease and type of hormone replacement therapy (iCARE-Lit model for women 50 years or older) was set to missing for all subjects. Meta-analysis of the iCARE-Lit model for women younger than 50 years included GS, NHS II, and UK Biobank. Meta-analysis of the iCARE-Lit model for women 50 years or older additionally included CPS II, EPIC NL, EPIC UK, KARMA, MMHS, NHS, PLCO, and WGHS. The AUC estimates were adjusted for age at enrollment. Abbreviations: AUC = area under the curve,  $\chi^2$  = chi-square test statistic, CPS = Cancer Prevention Study, EPIC = European Prospective Investigation into Cancer and Nutrition, GS = Generations Study, iCARE-Lit = iCARE model based on literature review, KARMA = KARolinska MAMmography Project, MMHS = Mayo Mammography Health Study, NHS = Nurses' Health Study, PLCO = Prostate, Lung, Colorectal, Ovarian Cancer Screening Trial, PRS = polygenic risk score, UK = United Kingdom, WGHS = Women's Genome Health Study.



**Figure 3.** Cohort-specific AUCs for the integrated iCARE-Lit models (with age at enrollment, classical risk factors and PRS). Classical risk factors include age at menarche, age at first live birth, parity, oral contraceptive use, age at menopause, hormone replacement therapy use, type of hormone replacement therapy, alcohol intake, height, BMI, breast cancer family history (i.e., presence or absence of breast cancer in at least one first degree relative), and benign breast disease. The vertical dashed line and gray region represent the meta-analyzed AUC estimate and the 95% confidence interval for the meta-analyzed AUC. Abbreviations: AU = Australia, AUC = area under the curve, CPS = Cancer Prevention Study, DE = Germany, EPIC = European Prospective Investigation into Cancer and Nutrition, GS = Generations Study, KARMA = KARolinska MAMmography Project, MCCS = Melbourne Collaborative Cohort Study, MMHS = Mayo Mammography Health Study, NHS = Nurses' Health Study, NL = the Netherlands, PRS = polygenic risk score, PLCO = Prostate, Lung, Colorectal, Ovarian Cancer Screening Trial, PROCAS = Predicting Risk Of Breast CANcer at Screening, SE = Sweden, UK = United Kingdom, US = United States, WGHS = Women's Genome Health Study.

**Figure 4.** Absolute risk calibration for the integrated iCARE-Lit model with classical risk factors and PRS for women younger than 50 years (4 cohorts). Risk categories were defined based on deciles of predicted five-year absolute risk. Classical risk factors include age at menarche, age at first live birth, parity, oral contraceptive use, age at menopause, hormone replacement therapy use, type of hormone replacement therapy, alcohol intake, height, BMI, breast cancer family history (i.e., presence or absence of breast cancer in at least one first degree relative), and benign breast disease. Abbreviations: E = Average of predicted five-year risk in the highest decile of predicted five-year risk, GS = Generations Study, NHS = Nurses' Health Study, O = observed proportion of subjects developing breast cancer in five years in the highest decile of predicted five-year risk, UK = United Kingdom.

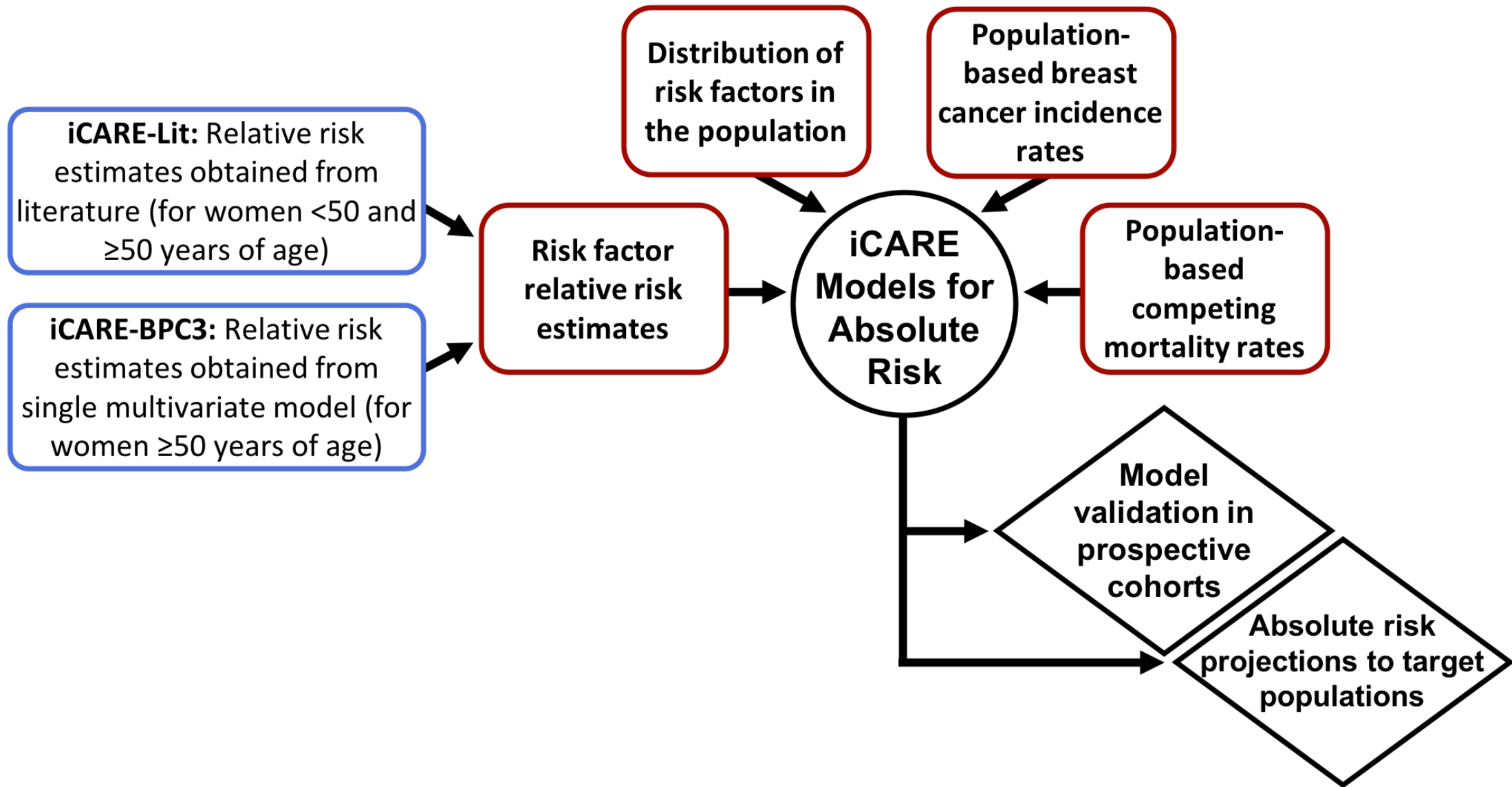
**Figure 5.** Absolute risk calibration for the integrated iCARE-Lit model with classical risk factors and PRS for women 50 years or older (15 cohorts). Risk categories were defined based on deciles of predicted five-year absolute risk. Classical risk factors include age at menarche, age at first live birth, parity, oral contraceptive use, age at menopause, hormone replacement therapy use, type of hormone replacement therapy, alcohol intake, height, BMI, breast cancer family history (i.e., presence or absence of breast cancer in at least one first degree relative), and benign breast disease. Abbreviations: CPS = Cancer Prevention Study, DE = Germany, E = Average of predicted five-year risk in the highest decile of predicted five-year risk, EPIC = European Prospective Investigation into Cancer and Nutrition, GS = Generations Study, KARMA = KARolinska MAMmography Project, MCCS = Melbourne Collaborative Cohort Study, MMHS = Mayo Mammography Health Study, NHS = Nurses' Health Study, NL = the Netherlands, O = observed proportion of subjects developing breast cancer in five years in the highest decile of predicted five-year risk, PLCO = Prostate, Lung, Colorectal, Ovarian Cancer Screening Trial, PROCAS = Predicting Risk Of Breast CAncer at Screening, UK = United Kingdom, WGHS = Women's Genome Health Study.

**Figure 6.** Women of European ancestry aged 50-70 years in the general populations of the six countries (Australia, Germany, The Netherlands, Sweden, the UK, the US) expected to be identified at low and high risk of breast cancer according to two risk thresholds and the incident cases of breast cancer expected to occur in these groups within a five-year interval. The expected number of women is calculated using 2017 population estimates (N = 2,960,506) from Australian Bureau of Statistics for Australia, 2016 population estimates (N=12,024,487) from the Federal Statistical Office for Germany, 2016 population estimates (N = 2,356,691) from the Central Agency for Statistics for the Netherlands, 2016 population estimates (N=1,249,695) from Statistics Sweden for Sweden, mid-2016 population estimates (N = 8,275,453) from the Office of National Statistics for the UK and mid-2016 population estimates (N = 30,030,821) from US Census Bureau for the US. The expected numbers of cases are estimated using the average predicted five-year risk in each population, calculated using the country-

specific breast cancer incidence rates and risk factor distributions (Table S4). The 1.13% risk threshold corresponds to the average five-year risk of US women aged 50 years. The 3% threshold is used by US Preventive Services Task Force for recommending risk reducing medications. Classical risk factors correspond to the iCARE-Lit model and include age at menarche, age at first live birth, parity, oral contraceptive use, age at menopause, hormone replacement therapy use, type of hormone replacement therapy, alcohol intake, height, BMI, breast cancer family history (i.e., presence or absence of breast cancer in at least one first degree relative), and benign breast disease. Abbreviations: AR = Absolute risk, PRS = polygenic risk score, UK = United Kingdom, US = United States.

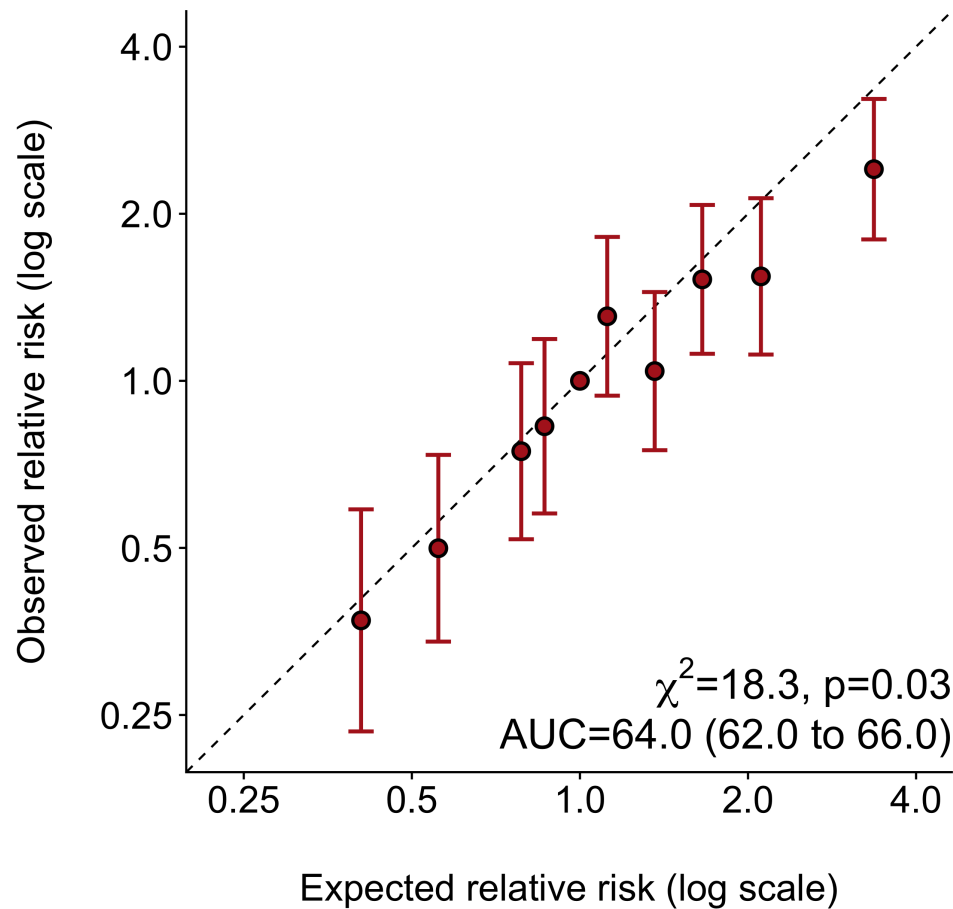
**Table 1.** Reclassification of women at high-risk thresholds after incorporating 313-SNP PRS to the classical risk factors. We report the number of women and future cases (i.e., women expected to develop breast cancer within five-years) above or below the risk threshold based on the classical risk factors and the number and percentage of these women moving below the threshold (down), above the threshold (up) and a total number and percentage of re-classified women. Classical risk factors correspond to the iCARE-Lit model and include age at menarche, age at first live birth, parity, oral contraceptive use, age at menopause, hormone replacement therapy use, type of hormone replacement therapy, alcohol intake, height, BMI, breast cancer family history (i.e., presence or absence of breast cancer in at least one first degree relative), and benign breast disease. The 3% cutoff corresponds to the US Preventive Services Task Force recommendation for risk-lowering drugs and the 6% cutoff corresponds to a cutoff for very high risk used in the WISDOM trial. Abbreviations: AR = absolute risk, PRS = polygenic risk score, SNP = single nucleotide polymorphism.

	UK population				US population			
	3% risk		6% risk		3% risk		6% risk	
	Total women	Future cases	Total women	Future cases	Total women	Future cases	Total women	Future cases
<b>Based on classical risk factors only</b>								
Above threshold, n (%)	940,889 (11.4)	34,955 (21.1)	36,871 (0.4)	2,407 (1.5)	4,224,349 (14.1)	169,055 (27.2)	223,792 (0.7)	16,327 (2.6)
Below threshold, n (%)	7,334,564 (88.6)	130,768 (78.9)	8,238,582 (99.6)	163,316 (98.5)	25,806,471 (85.9)	452,003 (72.8)	29,807,028 (99.3)	604,731 (97.4)
Total, N	8,275,453	165,723	8,275,453	165,723	30,030,820	621,058	30,030,820	621,058
<b>Based on classical risk factors + PRS</b>								
Reclassified, n (%)	1,265,899 (15.3)	42,929 (25.9)	140,656 (1.7)	9,970 (6.0)	4,424,697 (14.7)	148,710 (23.9)	790,733 (2.6)	55,387 (8.9)
Moving down, n (%)	402,848 (4.9)	8,462 (5.1)	17,753 (0.2)	771 (0.5)	1,660,327 (5.5)	36,180 (5.8)	108,699 (0.3)	4,218 (0.7)
Moving up, n (%)	863,051 (10.4)	34,467 (20.8)	122,903 (1.5)	9,199 (5.5)	2,764,370 (9.2)	112,530 (18.1)	682,034 (2.3)	51,169 (8.2)



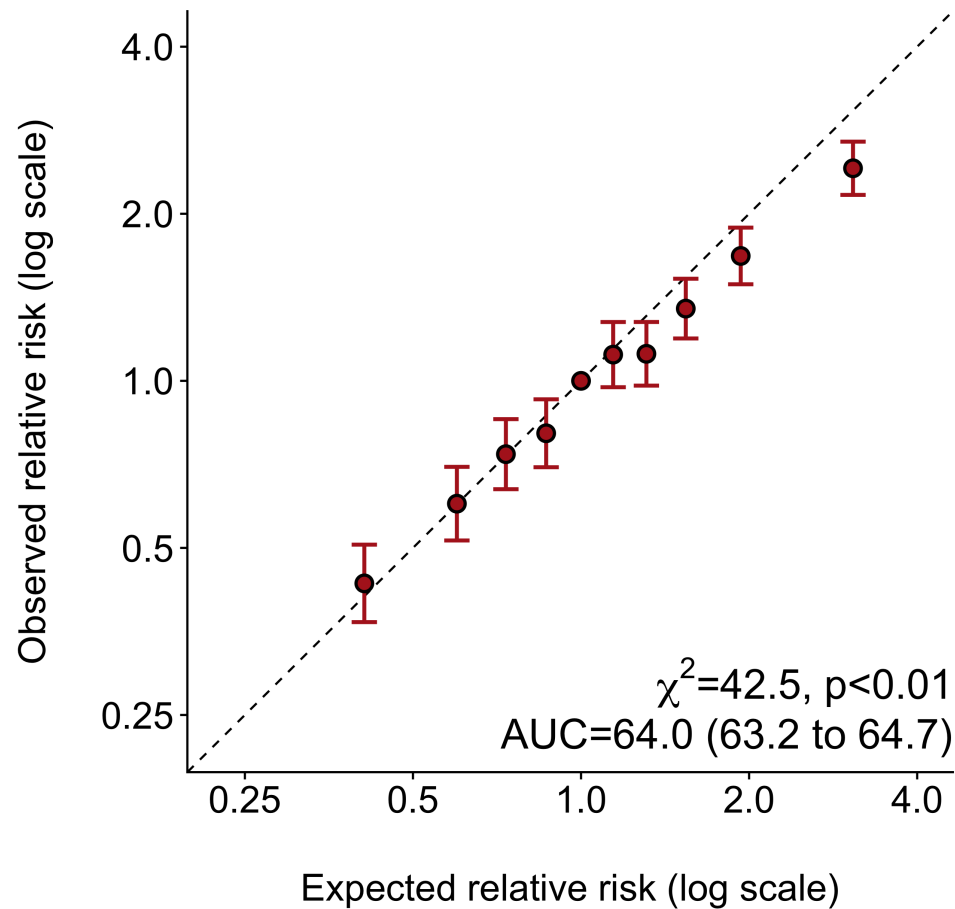
## iCARE-Lit Model (ages <50)

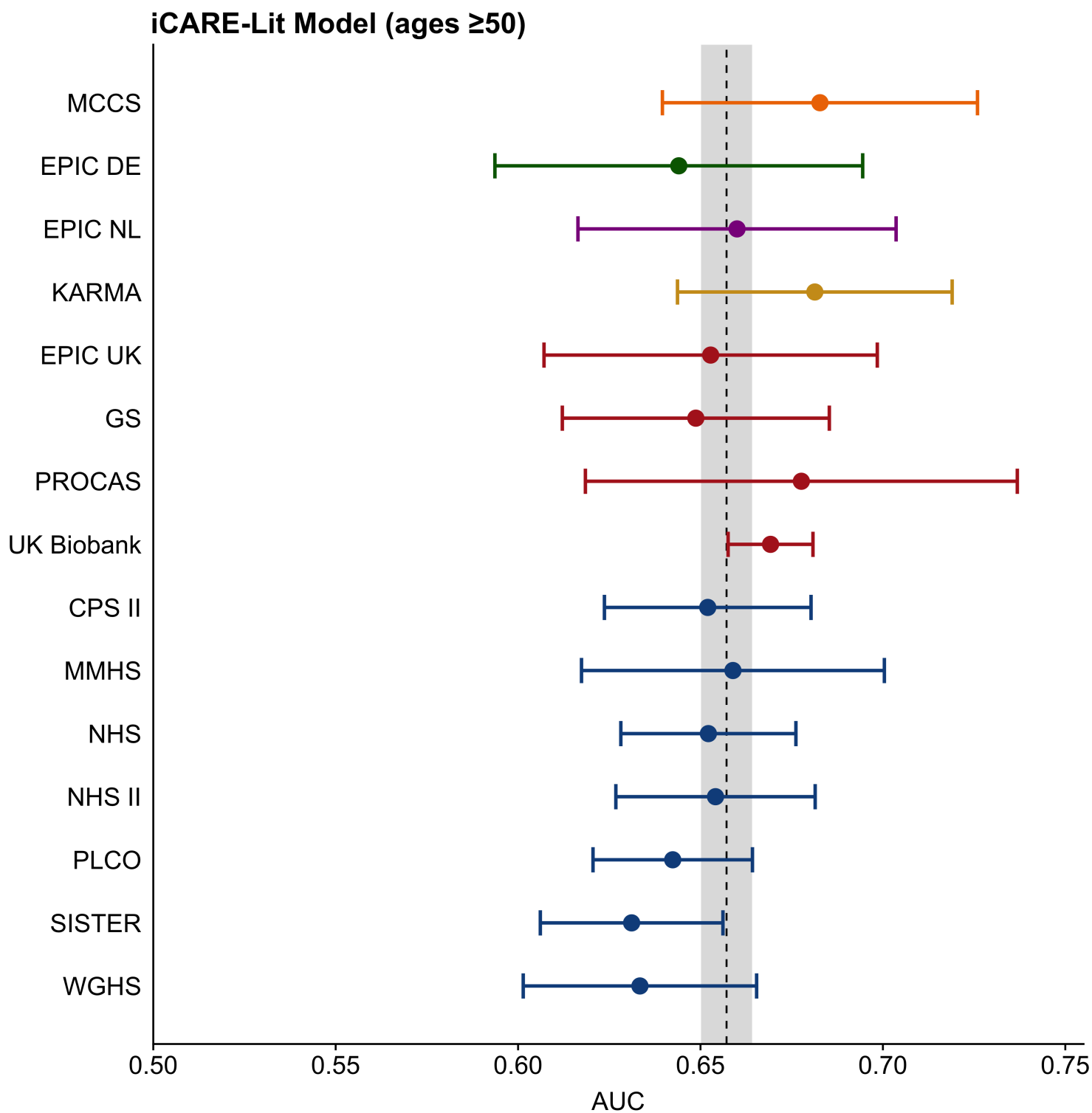
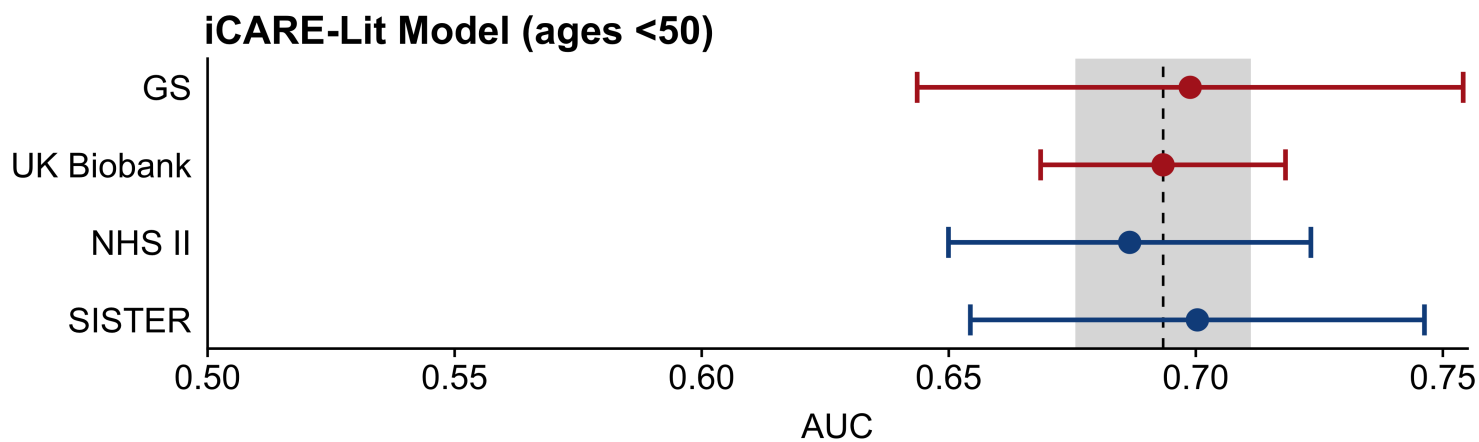
891 cases, 54,058 controls



## iCARE-Lit Model (ages $\geq 50$ )

5,920 cases, 232,743 controls

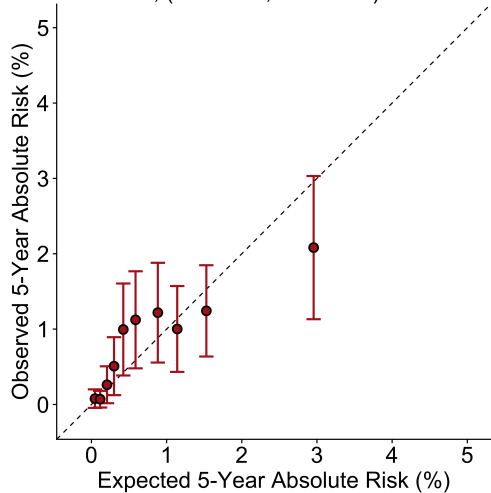




Median (range) of E/O in the highest risk decile = 0.9 (0.7 to 1.4)

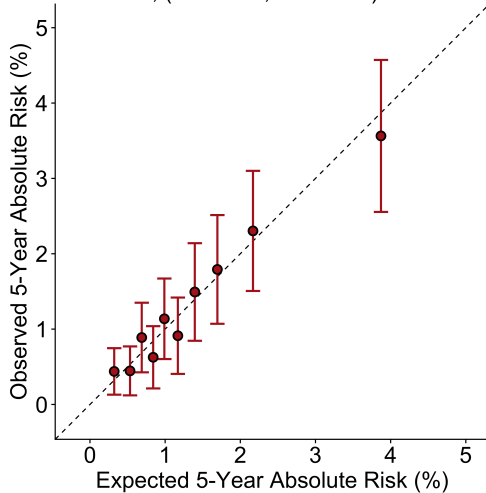
### GS

Highest decile:  
E/O = 1.4, (E = 3.0%, O = 2.1%)



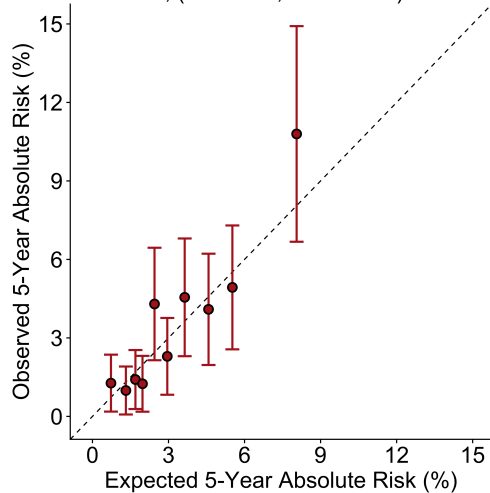
### NHS II

Highest decile:  
E/O = 1.1, (E = 3.9%, O = 3.6%)



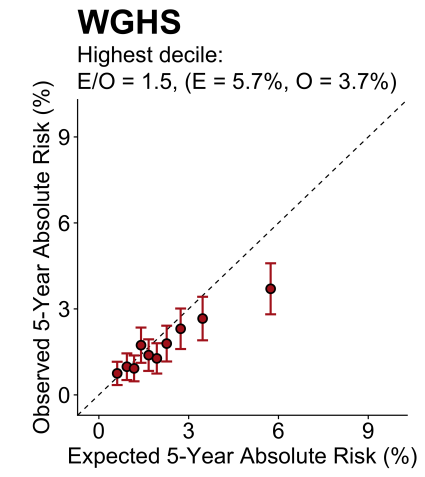
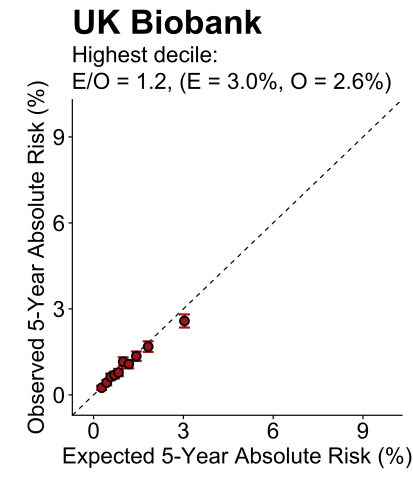
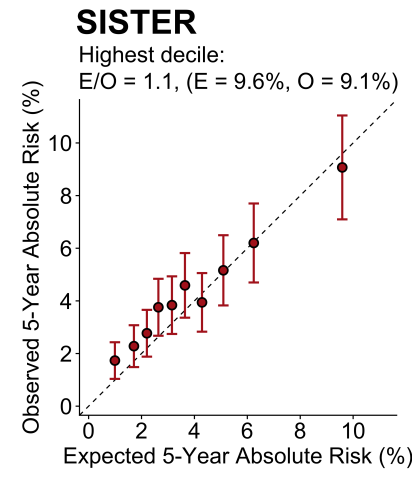
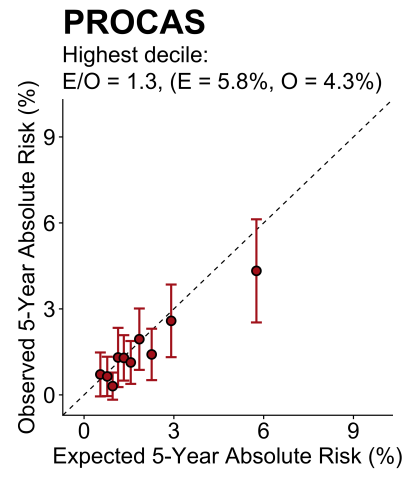
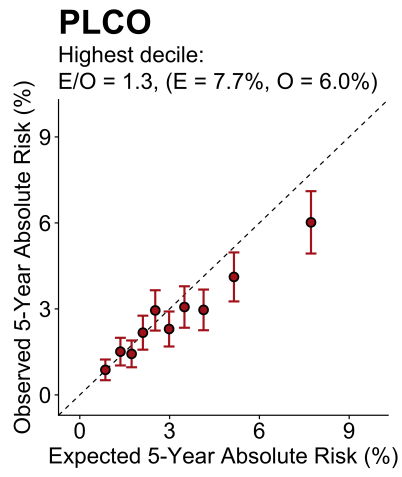
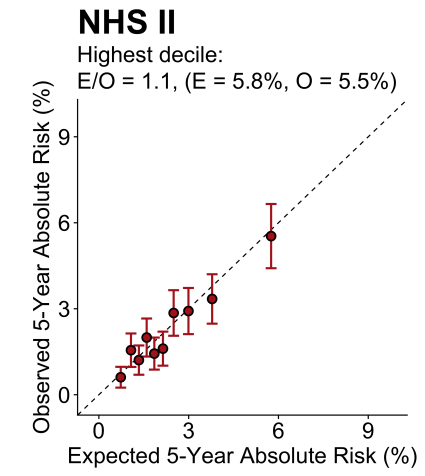
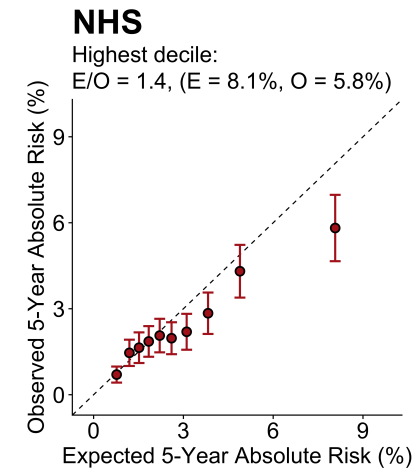
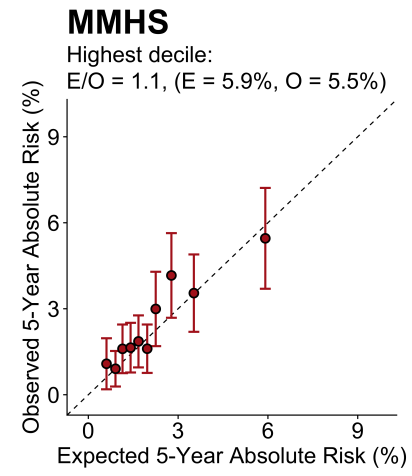
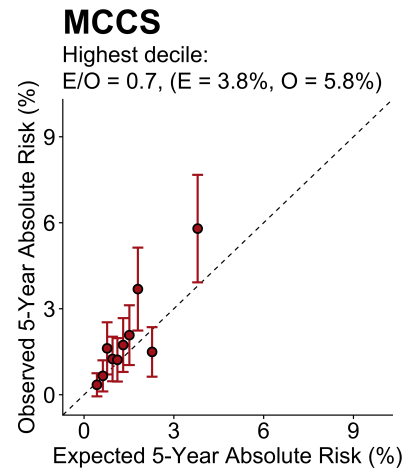
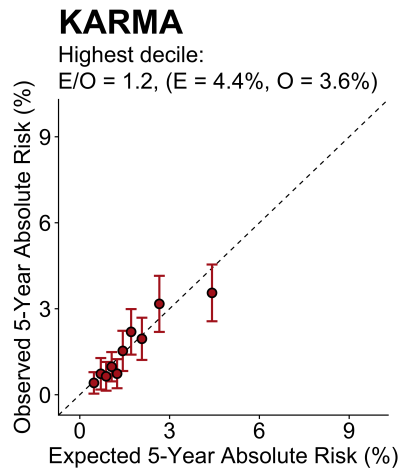
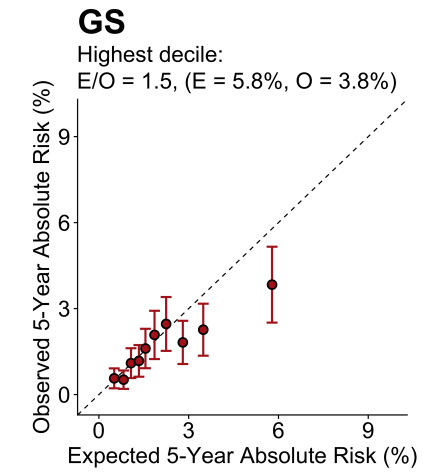
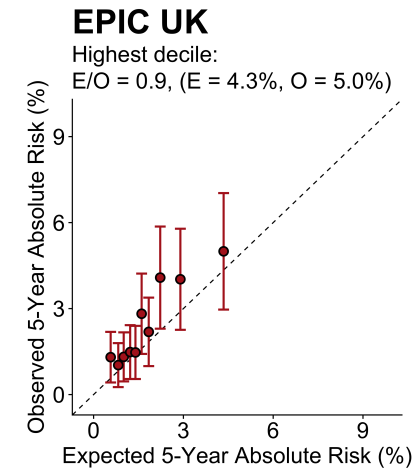
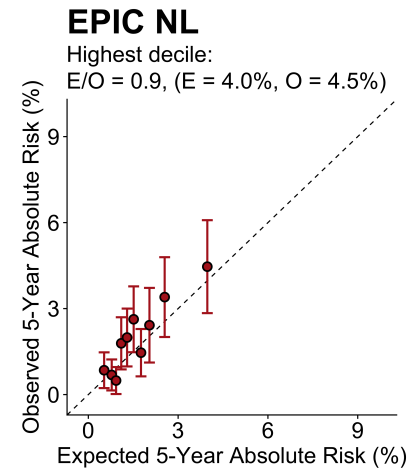
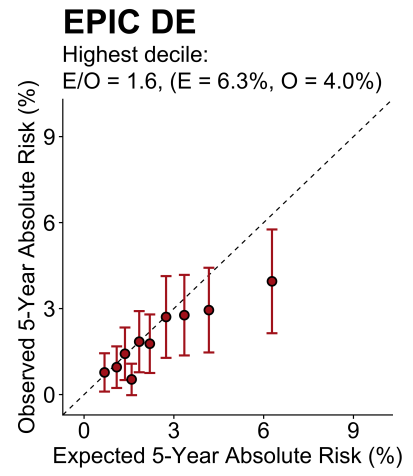
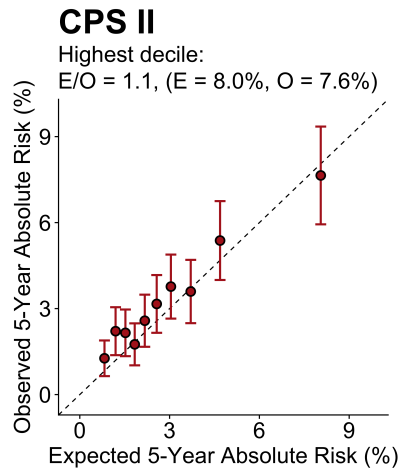
### SISTER

Highest decile:  
E/O = 0.7, (E = 8.1%, O = 10.8%)

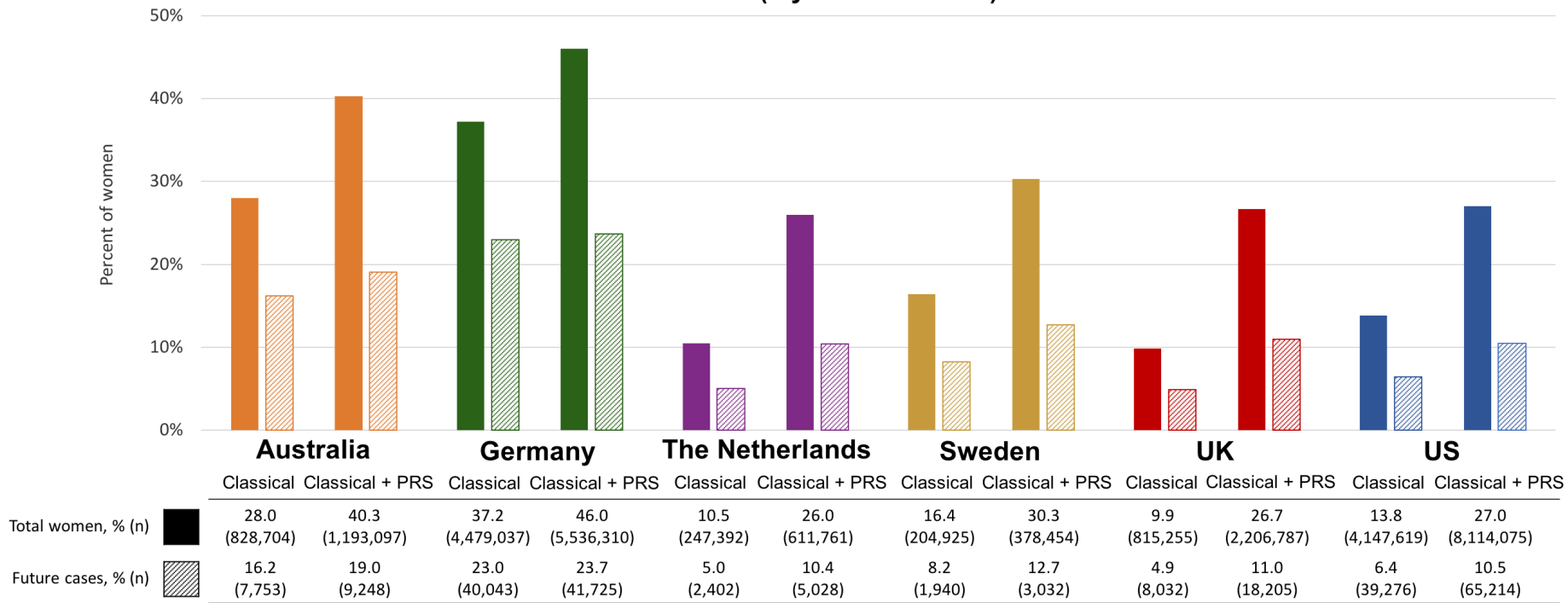




Median (range) of E/O in the highest risk decile = 1.2 (0.7 to 1.6)



### Low risk (5-year AR $\leq 1.13\%$ )



### High risk (5-year AR $> 3\%$ )

