

Differences in polygenic score distributions in European ancestry populations: implications for breast cancer risk prediction

Kristia Yiangou¹, Nasim Mavaddat², Joe Dennis², Maria Zanti¹, Qin Wang², Manjeet K. Bolla², Mustapha Abubakar³, Thomas U. Ahearn³, Irene L. Andrulis^{4,5}, Hoda Anton-Culver⁶, Natalia N. Antonenkova⁷, Volker Arndt⁸, Kristan J. Aronson⁹, Annelie Augustinsson¹⁰, Adinda Baten¹¹, Sabine Behrens¹², Marina Bermisheva^{13,14}, Amy Berrington de Gonzalez¹⁵, Katarzyna Białkowska¹⁶, Nicholas Boddicker¹⁷, Clara Bodelon¹⁸, Natalia V. Bogdanova^{7,19,20}, Stig E. Bojesen²¹⁻²³, Kristen D. Brantley²⁴, Hiltrud Brauch²⁵⁻²⁷, Hermann Brenner^{8,28,29}, Nicola J. Camp³⁰, Federico Canzian³¹, Jose E. Castelao³², Melissa H. Cessna^{33,34}, Jenny Chang-Claude^{12,35}, Georgia Chenevix-Trench³⁶, Wendy K. Chung³⁷, NBCS Collaborators³⁸⁻⁴⁹, Sarah V. Colonna³⁰, Fergus J. Couch⁵⁰, Angela Cox⁵¹, Simon S. Cross⁵², Kamila Czene⁵³, Mary B. Daly⁵⁴, Peter Devilee^{55,56}, Thilo Dörk²⁰, Alison M. Dunning⁵⁷, Diana M. Eccles⁵⁸, A. Heather Eliassen^{24,59,60}, Christoph Engel^{61,62}, Mikael Eriksson⁵³, D. Gareth Evans^{63,64}, Peter A. Fasching⁶⁵, Olivia Fletcher⁶⁶, Henrik Flyger⁶⁷, Lin Fritschi⁶⁸, Manuela Gago-Dominguez⁶⁹, Aleksandra Gentry-Maharaj^{70,71}, Anna González-Neira^{72,73}, Pascal Guénel⁷⁴, Eric Hahnen^{75,76}, Christopher A. Haiman⁷⁷, Ute Hamann⁷⁸, Jaana M. Hartikainen^{79,80}, Vikki Ho⁸¹, James Hodge¹⁸, Antoinette Hollestelle⁸², Ellen Honisch⁸³, Maartje J. Hooning⁸², Reiner Hoppe^{25,84}, John L. Hopper⁸⁵, Sacha Howell⁸⁶⁻⁸⁸, Anthony Howell⁸⁹, ABCTB Investigators⁹⁰, kConFab Investigators^{91,92}, Simona Jakovchevska⁹³, Anna Jakubowska^{16,94}, Helena Jernström¹⁰, Nichola Johnson⁶⁶, Rudolf Kaaks¹², Elza K. Khusnutdinova^{13,95}, Cari M. Kitahara⁹⁶, Stella Koutros³, Vessela N. Kristensen^{39,49}, James V. Lacey^{97,98}, Diether Lambrechts^{99,100}, Flavio Lejbkowitz¹⁰¹, Annika Lindblom^{102,103}, Michael Lush², Arto Mannermaa^{80,104,105}, Dimitrios Mavroudis¹⁰⁶, Usha Menon⁷⁰, Rachel A. Murphy^{107,108}, Heli Nevanlinna¹⁰⁹, Nadia Obi^{110,111}, Kenneth Offit^{112,113}, Tjoung-Won Park-Simon²⁰, Alpa V. Patel¹⁸, Cheng Peng⁵⁹, Paolo Peterlongo¹¹⁴, Guillermo Pita⁷², Dijana Plaseska-Karanfilska⁹³, Katri Pylkäs^{115,116}, Paolo Radice¹¹⁷, Muhammad U. Rashid^{78,118}, Gad Rennert¹¹⁹, Eleanor Roberts⁸⁶, Juan Rodriguez⁵³, Atocha Romero¹²⁰, Efraim H. Rosenberg¹²¹, Emmanouil Saloustros¹²², Dale P. Sandler¹²³, Elinor J. Sawyer¹²⁴, Rita K. Schmutzler^{75,76,125}, Christopher G. Scott¹⁷, Xiao-Ou Shu¹²⁶, Melissa C. Southey¹²⁷⁻¹²⁹, Jennifer Stone^{85,130}, Jack A. Taylor^{123,131}, Lauren R. Teras¹⁸, Irma van de Beek¹³², Walter Willett^{24,59,60}, Robert Winqvist^{115,116}, Wei Zheng¹²⁶, Celine M. Vachon¹³³, Marjanka K. Schmidt¹³⁴⁻¹³⁶, Per Hall^{53,137}, Robert J. MacInnis^{85,129}, Roger L. Milne^{85,127,129}, Paul D.P. Pharoah¹³⁸, Jacques Simard¹³⁹, Antonis C. Antoniou², Douglas F. Easton^{2,57}, Kyriaki Michailidou^{1,2*}

1. Biostatistics Unit, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus, 2371.
2. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK, CB1 8RN.
3. Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA, 20850.
4. Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada, M5G 1X5.
5. Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada, M5S 1A8.
6. Department of Medicine, Genetic Epidemiology Research Institute, University of California Irvine, Irvine, CA, USA, 92617.
7. NN Alexandrov Research Institute of Oncology and Medical Radiology, Minsk, Belarus, 223040.
8. Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany, 69120.
9. Department of Public Health Sciences, and Cancer Research Institute, Queen's University, Kingston, ON, Canada, K7L 3N6.
10. Oncology, Clinical Sciences in Lund, Lund University, Lund, Sweden, 221 85.
11. Leuven Multidisciplinary Breast Center, Department of Oncology, Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium, 3000.
12. Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, 69120.
13. Institute of Biochemistry and Genetics of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia, 450054.
14. St Petersburg State University, St, Petersburg, Russia, 199034.
15. Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK, SM2 5NG.
16. Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland, 71-252.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

17. Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA, 55905.
18. Department of Population Science, American Cancer Society, Atlanta, GA, USA, 30303.
19. Department of Radiation Oncology, Hannover Medical School, Hannover, Germany, 30625.
20. Gynaecology Research Unit, Hannover Medical School, Hannover, Germany, 30625.
21. Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark, 2730.
22. Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark, 2730.
23. Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, 2200.
24. Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA, USA, 02115.
25. Dr Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany, 70376.
26. iFIT-Cluster of Excellence, University of Tübingen, Tübingen, Germany, 72074.
27. German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Partner Site Tübingen, Tübingen, Germany, 72074.
28. Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany, 69120.
29. German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany, 69120.
30. Department of Internal Medicine and Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA, 84112.
31. Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany, 69120.
32. Oncology and Genetics Unit, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS) Foundation, Complejo Hospitalario Universitario de Santiago, SERGAS, Vigo, Spain, 36312.
33. Department of Pathology, Intermountain Healthcare, Salt Lake City, UT, USA, 84143.
34. Intermountain Biorepository, Intermountain Healthcare, Salt Lake City, UT, USA, 84143.
35. Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany, 20246.
36. Cancer Research Program, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia, 4006.
37. Departments of Pediatrics and Medicine, Columbia University, New York, NY, USA, 10032.
38. Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital-Radiumhospitalet, Oslo, Norway, 0379.
39. Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway, 0450.
40. Department of Research, Vestre Viken Hospital, Drammen, Norway, 3019.
41. Section for Breast- and Endocrine Surgery, Department of Cancer, Division of Surgery, Cancer and Transplantation Medicine, Oslo University Hospital-Ullevål, Oslo, Norway, 0450.
42. Department of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway, 0379.
43. Department of Pathology, Akershus University Hospital, Lørenskog, Norway, 1478.
44. Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway, 0379.
45. Department of Oncology, Division of Surgery, Cancer and Transplantation Medicine, Oslo University Hospital-Radiumhospitalet, Oslo, Norway, 0379.
46. National Advisory Unit on Late Effects after Cancer Treatment, Oslo University Hospital, Oslo, Norway, 0379.
47. Department of Oncology, Akershus University Hospital, Lørenskog, Norway, 1478.
48. Oslo Breast Cancer Research Consortium, Oslo University Hospital, Oslo, Norway, 0379.
49. Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway, 0379.
50. Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA, 55905.
51. Division of Clinical Medicine, School of Medicine and Population Health, University of Sheffield, Sheffield, UK, S10 2TN.
52. Division of Neuroscience, School of Medicine and Population Health, University of Sheffield, Sheffield, UK, S10 2TN.
53. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, 171 65.
54. Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA, USA, 19111.
55. Department of Pathology, Leiden University Medical Center, Leiden, the Netherlands, 2333 ZA.
56. Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands, 2333 ZA.
57. Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK, CB1 8RN.
58. Faculty of Medicine, University of Southampton, Southampton, UK, SO17 1BJ.
59. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA, 02115.
60. Department of Nutrition, Harvard TH Chan School of Public Health, Boston, MA, USA, 02115.
61. Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany, 04107.
62. LIFE - Leipzig Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Germany, 04103.

63. Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK, M13 9WL.
64. North West Genomics Laboratory Hub, Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK, M13 9WL.
65. Department of Gynecology and Obstetrics, Comprehensive Cancer Center Erlangen-EMN, Friedrich-Alexander University Erlangen-Nuremberg, University Hospital Erlangen, Erlangen, Germany, 91054.
66. The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK, SW7 3RP.
67. Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark, 2730.
68. School of Population Health, Curtin University, Perth, Western Australia, Australia, 6102.
69. Cancer Genetics and Epidemiology Group, Genomic Medicine Group, Fundación Instituto de Investigación Sanitaria de Santiago de Compostela (FIDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain, 15706.
70. MRC Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, London, UK, WC1V 6LJ.
71. Department of Women's Cancer, Elizabeth Garrett Anderson Institute for Women's Health, University College London, London, UK.
72. Human Genotyping Unit-CeGen, Spanish National Cancer Research Centre (CNIO), Madrid, Spain, 28029.
73. Spanish Network on Rare Diseases (CIBERER).
74. Team 'Exposome and Heredity', CESP, Gustave Roussy, INSERM, University Paris-Saclay, UVSQ, Villejuif, France, 94805.
75. Center for Familial Breast and Ovarian Cancer, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany, 50937.
76. Center for Integrated Oncology (CIO), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany, 50937.
77. Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA, 90033.
78. Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany, 69120.
79. Cancer RC, University of Eastern Finland, Kuopio, Finland, 70210.
80. Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland, 70210.
81. Health Innovation and Evaluation Hub, Université de Montréal Hospital Research Centre (CRCHUM), Montréal, Québec, Canada.
82. Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands, 3015 GD.
83. Department of Gynecology and Obstetrics, University Hospital Düsseldorf, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany, 40225.
84. University of Tübingen, Tübingen, Germany, 72074.
85. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia, 3010.
86. Division of Cancer Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK.
87. Nightingale/Prevent Breast Cancer Centre, Wythenshawe Hospital, Manchester University NHS Foundation Trust, Manchester, UK.
88. Manchester Breast Centre, Manchester Cancer Research Centre, The Christie Hospital, Manchester, UK.
89. Division of Cancer Sciences, University of Manchester, Manchester, UK, M13 9PL.
90. Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia, 2145.
91. Research Department, Peter MacCallum Cancer Center, Melbourne, Victoria, Australia, 3000.
92. Sir Peter MacCallum Department of Oncology, The University of Melbourne, Parkville, Victoria, Australia, 3000.
93. Research Centre for Genetic Engineering and Biotechnology 'Georgi D, Efremov', MASA, Skopje, Republic of North Macedonia, 1000.
94. Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin, Poland, 171-252.
95. Department of Genetics and Fundamental Medicine, Ufa University of Science and Technology, Ufa, Russia, 450076.
96. Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD, USA, 20892.
97. Department of Computational and Quantitative Medicine, City of Hope, Duarte, CA, USA, 91010.
98. City of Hope Comprehensive Cancer Center, City of Hope, Duarte, CA, USA, 91010.
99. Laboratory for Translational Genetics, Department of Human Genetics, KU Leuven, Leuven, Belgium, 3000.

100. VIB Center for Cancer Biology, VIB, Leuven, Belgium, 3001.
101. Carmel Medical Center, Haifa, Israel.
102. Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden, 171 76.
103. Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden, 171 76.
104. Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland, 70210.
105. Biobank of Eastern Finland, Kuopio University Hospital, Kuopio, Finland.
106. Department of Medical Oncology, University Hospital of Heraklion, Heraklion, Greece, 711 10.
107. School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada, V6T 1Z4.
108. Cancer Control Research, BC Cancer Agency, Vancouver, BC, Canada, V5Z 1L3.
109. Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland, 00290.
110. Institute for Occupational and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, 20246.
111. Institute for Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, 20246.
112. Clinical Genetics Research Lab, Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY, USA, 10065.
113. Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA, 10065.
114. Genome Diagnostics Program, IFOM ETS - the AIRC Institute of Molecular Oncology, Milan, Italy, 20139.
115. Laboratory of Cancer Genetics and Tumor Biology, Translational Medicine Research Unit, Biocenter Oulu, University of Oulu, Oulu, Finland, 90220.
116. Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre Oulu, Oulu, Finland, 90220.
117. Unit of Predictive Medicine, Molecular Bases of Genetic Risk, Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy, 20133.
118. Department of Basic Sciences, Shaikat Khanum Memorial Cancer Hospital and Research Centre (SKMCH & RC), Lahore, Pakistan, 54000.
119. Technion, Faculty of Medicine and Association for Promotion of Research in Precision Medicine, Haifa, Israel.
120. Medical Oncology Department, Hospital Universitario Puerta de Hierro, Madrid, Spain, 28222.
121. Department of Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, the Netherlands, 1066 CX.
122. Department of Oncology, University Hospital of Larissa, Larissa, Greece, 411 10.
123. Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA, 27709.
124. School of Cancer & Pharmaceutical Sciences, Comprehensive Cancer Centre, Guy's Campus, King's College London, London, UK.
125. Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany, 50931.
126. Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA, 37232.
127. Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia, 3168.
128. Department of Clinical Pathology, The University of Melbourne, Melbourne, Victoria, Australia, 3010.
129. Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia, 3004.
130. Genetic Epidemiology Group, School of Population and Global Health, University of Western Australia, Perth, Western Australia, Australia, 6000.
131. Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA, 27709.
132. Department of Clinical Genetics, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, the Netherlands, 1066 CX.
133. Department of Quantitative Health Sciences, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA, 55905.
134. Division of Molecular Pathology, The Netherlands Cancer Institute, Amsterdam, the Netherlands, 1066 CX.
135. Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, the Netherlands, 1066 CX.
136. Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands, 2333 ZA.
137. Department of Oncology, Södersjukhuset, Stockholm, Sweden, 118 83.
138. Department of Computational Biomedicine, Cedars-Sinai Medical Center, West Hollywood, CA, USA, 90069.

139. Genomics Center, Centre Hospitalier Universitaire de Québec – Université Laval Research Center, Québec City, Québec, Canada, G1V 4G2.

*corresponding author; Kyriaki Michailidou, email: kyriakimi@cing.ac.cy

Abstract

The 313-variant polygenic risk score (PRS₃₁₃) provides a promising tool for breast cancer risk prediction. However, evaluation of the PRS₃₁₃ across different European populations which could influence risk estimation has not been performed. Here, we explored the distribution of PRS₃₁₃ across European populations using genotype data from 94,072 females without breast cancer, of European-ancestry from 21 countries participating in the Breast Cancer Association Consortium (BCAC) and 225,105 female participants from the UK Biobank. The mean PRS₃₁₃ differed markedly across European countries, being highest in south-eastern Europe and lowest in north-western Europe. Using the overall European PRS₃₁₃ distribution to categorise individuals leads to overestimation and underestimation of risk in some individuals from south-eastern and north-western countries, respectively. Adjustment for principal components explained most of the observed heterogeneity in mean PRS. Country-specific PRS distributions may be used to calibrate risk categories in individuals from different countries.

Introduction

Genetic susceptibility to breast cancer is influenced by multiple genetic variants which contribute different levels of risk to the disease (1-6). Genome-wide Association Studies (GWAS) have identified thus far a large number of common, low-risk variants that each contribute a small risk to the disease but can be combined into Polygenic Risk Scores (PRSs) with larger effect (7, 8). PRSs provide a promising tool for clinical risk prediction of breast cancer by stratifying women into different categories of breast cancer risk (9-11), and may be used to inform targeted screening and prevention strategies (12-20).

Mavaddat et al., (2019) constructed a 313-variant PRS (PRS₃₁₃) for breast cancer, using data for women of European ancestry from the Breast Cancer Association Consortium (BCAC) (11). In prospective validation studies, this PRS was estimated to be associated with a relative risk for breast cancer of approximately 1.6 per standard deviation increase, and its discriminatory ability, measured in terms of area under the ROC curve (AUC), was 0.63. The lifetime absolute risk of developing breast cancer for individuals in the lowest percentile of the PRS₃₁₃ risk distribution was estimated to be ~2%, while for those in the highest percentile it was ~33%. PRS₃₁₃ has been incorporated into the multifactorial BOADICEA (Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm) model which is available via the CanRisk tool (14, 21, 22) (www.canrisk.org) and, together with other lifestyle and genetic risk factors, has been shown to improve risk

stratification in European and European ancestry populations (14, 23-27). PRS₃₁₃ has also been shown to be transferrable to women of other ethnic backgrounds, although the strength of the association with breast cancer risk was attenuated compared with that for women of European ancestry (OR per SD (95% CI) 1.52 (1.49-1.56), AUC = 0.61 in women of east Asian ancestry; OR 1.27 (1.23-1.31), AUC = 0.57 in women of African ancestry (28-30)).

Although several studies have investigated the transferability of PRS developed in European ancestry populations to non-European populations (31-34), the PRS distributions across different European countries has not been extensively evaluated. Differences in the PRS distribution, if not appropriately accounted for, could lead to inappropriate risk classification, with implications for clinical management.

In this study, we aimed to examine the distribution of the PRS₃₁₃ across 17 countries in Europe, together with individuals of European ancestry from Australia, Canada, Israel and the USA. Similar analyses were performed using data from the UK Biobank, stratifying individuals by country of birth. We explored different approaches to account for the differences in the distribution, and investigated the implications of distribution differences across countries in breast cancer risk prediction.

Materials and methods

Study populations and Genotyping

Breast Cancer Association Consortium dataset

The BCAC dataset used here consisted of 110,260 female invasive breast cancer cases and 94,072 female healthy controls of European ancestry, recruited into 84 studies from 21 countries participating in the BCAC (**Supplementary Table 1A**). For simplicity and with attempt to explore the effect on the general female population, only the control data were used in these analyses as the distribution of the PRS in cases might vary between studies due to differences in study design (in particular oversampling of cases with a family history of disease). Samples from participating individuals were genotyped using the iCOGS (1) or OncoArray (3, 35) genotyping array. For samples genotyped using both arrays, the OncoArray genotype data were used. The iCOGS and the OncoArray datasets were imputed separately in a two-step manner using SHAPEIT (36) for phasing and IMPUTE2 for imputation. The Phase 3 (October 2014) release of the 1000 Genomes data (37) was used as the reference panel. More details on genotyping, quality control and imputation are given elsewhere (2, 3, 35). Ancestry-informative principal components (PCs), derived separately from the iCOGS and the OncoArray genotypes, were also calculated for all the samples, as previously described (3).

UK Biobank dataset

UK Biobank, is a prospective cohort study including more than 500,000 participants from England, Wales and Scotland, with age at recruitment between 40 to 69 years old, more details can be found elsewhere (38, 39). For

the analyses in this study, genotype data from females (genetic reported sex) participating in the UK Biobank were used. Individuals were excluded if they had a recorded breast cancer diagnosis (malignant neoplasm of breast or carcinoma in situ of breast) or had a personal history of malignant neoplasm of breast. Genetic ancestry was inferred using the FastPop software (40). Individuals self-reported “white” and with an estimated European ancestry proportion $\geq 80\%$ were retained in the analysis. Then, individuals were stratified by the “country of birth” field in the UK Biobank; only countries with at least 100 participants were used. After filtering, 225,105 females from 17 countries of Europe and from Australia, Canada, New Zealand, and the USA were used in the analyses (more details in **Supplementary Table 1B**). Samples were genotyped using the Affymetrix UK BiLEVE Axiom array and the Affymetrix UK Biobank Axiom array. Imputation data used were based on the Haplotype Reference Consortium (41), the UK10K +1000 Genomes panel references. More details on genotyping, quality control and imputation are given elsewhere (38). Ancestry informative PCs were also available (38).

All study participants gave written informed consent, and all the studies were approved by the relevant ethics committees. The use of the UK Biobank has been approved under application ID102655.

Statistical Analyses

PRS₃₁₃ was developed previously using a hard-thresholding stepwise forward regression approach, and included variants independently associated with breast cancer risk at a p-value cut off $< 10^{-5}$ (11). PRS₃₁₃ was calculated in each study participant using the following formula:

$$PRS_j = \beta_1 x_{j1} + \dots + \beta_k x_{jk} + \beta_{313} x_{j;313}$$

Where PRS_j is the PRS of individual j , x_{jk} is the estimated effect allele dosage for SNP_k carried by individual j and can take values between 0 and 2, and β_k is the weight for SNP_k in the PRS for overall breast cancer, as derived by Mavaddat et al. (11). PRS₃₁₃ was standardized to have unit SD in controls in the pooled dataset. Mavaddat et al. (11) also derived ER-specific versions of PRS₃₁₃, with weights optimised for predicting ER-positive or ER-negative breast cancer risk (**Supplementary Table 2**).

The main analyses focused on calculating the mean standardized PRS₃₁₃ in BCAC controls, using both the iCOGS and OncoArray datasets. These values were derived using linear regression with array type as a covariate and no intercept (so that estimates were generated for every country). Heterogeneity in the mean PRS₃₁₃ between countries was assessed using I^2 statistics and Q statistic p-values.

We also evaluated the distribution of the mean PRS by country of birth in female participants in the UK Biobank. Seven of the 313 variants were not available from the UK Biobank data and thus we used the remaining 306 variants in the analysis (PRS₃₀₆) (**Supplementary Table 2**). We also evaluated a “standard” breast cancer PRS available in the UK Biobank data, previously generated from external GWAS data (42), and was available for 224,776 individuals (**Supplementary Table 1B**).

Potential sources of the variability in the mean PRS₃₁₃ across the countries were explored in the BCAC dataset using three approaches. The PRS was first recalculated excluding variants in the *CHEK2* region. The protein truncating variant *CHEK2* c.1100delC is a relatively common founder variant that exhibits a large variation in frequency across Europe (43). Although it is not included in PRS₃₁₃, other variants in the PRS₃₁₃ are correlated with this variant. For this reason, the four variants in the *CHEK2* region included in PRS₃₁₃ (the *CHEK2* p. Ile157Thr variant, and variants at positions 29135543, 29203724 and 29551872 on chromosome 22, positions based on build 37) were removed, resulting in a 309-variant PRS (PRS₃₀₉). Mean and SE by country were recalculated for PRS₃₀₉, as described above.

Second, we examined the effect of removing variants with the most variable frequency across countries. For this analysis, the mean and SD of the effect allele frequency in controls of the pooled dataset was calculated for each of the 313 variants by country. Variants with a coefficient of variation (SD/mean) greater than 0.3 were removed. Means and SE of the newly constructed PRS were recalculated by country as described above.

Third, we explored the effect of adjusting for up to 10 ancestry-informative PCs, in addition to type of array. As the PCs derived from the iCOGS and OncoArray are not comparable, separate PCs for each were included in the regression. We explored the number of PCs that were required to eliminate the heterogeneity in the adjusted mean PRS₃₁₃, using the thresholds $I^2 < 10\%$ and $p\text{-value} > 0.05$. Similarly, for the UK Biobank dataset, PRS₃₀₆ was adjusted for up to 10 PCs, which were available in the UK Biobank.

As a complementary approach to generating population-specific estimates, we explored an empirical Bayes approach similar to that described by Clayton and Kaldor (44) for mapping disease rates. The motivation of this approach is that, if some of the variation in means among countries is genuine, while some is due to sampling variation, better estimates of the country-specific means can be obtained by “shrinking” the country-specific estimates towards the overall mean, by an amount depending on the sample size. In our implementation, we allowed the PRS means to be correlated between countries, using the autocorrelation matrix proposed in Clayton and Kaldor. A detailed description is given in **Supplementary Methods**.

To investigate the implications of PRS distribution differences in breast cancer risk prediction, we explored the proportion of women by country by percentile (<1%, 1%-5%, 5%-10%, 10%-20%, 20%-40%, 40%-60%, 60%-80%, 80%-90%, 90%-95%, 95%-99%, $\geq 99\%$ percentiles), based on the distribution cut-offs of either the full dataset or country-specific estimates. We also examined a specific risk estimation example using the CanRisk tool (14, 21, 22).

All analyses were performed in R (version 4.2.1) (45). Forest plots were generated using the metafor package (46). Maps were generated using the packages world map data from natural earth (rnatualearth) (47), the world vector map data from natural earth used in 'rnatualearth' (rnatualearthdata) (48), simple features for R (sf) (49) and interface to geometry engine (rgeos) (50).

Results

Geographic diversity in the mean PRS₃₁₃ across European ancestry populations

The mean PRS₃₁₃ in the BCAC controls differed markedly across European countries, with heterogeneity $I^2 = 80\%$ (p -value = 5.6×10^{-13}). The mean was highest in the Republic of North Macedonia (0.25), Greece (0.23), Russia (0.18) and Italy (0.12), and lowest in Ireland (-0.12). The mean estimates for Australia, Canada, Israel and the USA were close to the overall mean (**Figure 1; Figure 2; Table 1; Supplementary Table 3A**). A similar level of heterogeneity was observed for the ER-positive ($I^2 = 84\%$) and ER-negative PRS ($I^2 = 64\%$) (**Figure 2; Supplementary Table 3B**). There was no evidence of a difference in the SD of the PRS between countries. (**Supplementary Table 3A**).

The mean PRS₃₀₆ in female UK Biobank participants, stratified by country of birth, was also calculated (**Figure 3 and Supplementary Table 4**). There was strong evidence of heterogeneity in the PRS distribution ($I^2 = 66\%$, p -value = 2.3×10^{-06}). The pattern was generally similar to that seen in the BCAC dataset, with a higher PRS in individuals born in southern and eastern Europe (e.g. Cyprus, Russia, Italy) and lower in western Europe (e.g. Ireland). Similar results were found for the “standard” UK Biobank PRS ($I^2 = 87\%$, p -value = 1.4×10^{-25}) (**Figure 3 and Supplementary table 4**).

Exploring potential reasons for differences in mean standardized PRS between countries

Potential sources of the variability in the mean PRS₃₁₃ across the countries were explored in the BCAC dataset, using three approaches. We first evaluated the effect of removing variants in the *CHEK2* region on the distribution of the mean PRS₃₁₃ for the countries. After removing these four variants, the variation in the mean PRS₃₀₉ across countries in the controls remained similar that for PRS₃₁₃ ($I^2 = 83\%$, p -value = 9.4×10^{-16}). We next identified the variants with the most variable frequency from countries in the control dataset. Seventeen of the 313 variants had a coefficient of variation greater than 0.3 (**Supplementary Table 5**). Excluding these 17 variants did not reduced the variation in the mean PRS ($I^2 = 80\%$, p -value = 2.4×10^{-12}).

We next explored the effect of adjusting for PCs. When individuals in the BCAC dataset genotyped with OncoArray were plotted by the first two PCs, those from the same country separated clearly, in a pattern consistent with their geographical relationship (**Supplementary Figure 1**). This suggests that adjusting for PCs maybe an effective approach to reducing the variation in PRS distribution. When we adjusted the PRS for the leading PCs in the BCAC dataset, the I^2 reduced as each PC was added in the model and reached $< 10\%$ when adjusted for the first six PCs ($I^2 = 69\%$, 54% , 47% , 39% , 22% , 0% , and 0% when including 1, 2, 3, 4, 5, 6, and 10 PCs respectively) (**Table 1; Supplementary Table 3A; Supplementary Figure 2**). A similar result was obtained for the ER-positive PRS (**Supplementary Table 3B**), when adjusted for the first 6 PCs (Heterogeneity: $I^2 = 0\%$, p -value = 0.69). For the ER-negative PRS, however, the heterogeneity was not eliminated even when the PRS was adjusted for 10 PCs (Heterogeneity: $I^2 = 56\%$, p -value = 0.001) (**Supplementary Table 3B**). The predicted PRS of each individual, as derived from the fitted values of the

linear regression model of PRS adjusted for the first 6 PCs and array, were then used to calculate a predicted mean PRS₃₁₃ by country (**Table 1** and **Supplementary Table 3A**).

We repeated these analyses for PRS₃₀₆ using the UK Biobank dataset. I^2 reduced as each PC was added in the model and reached < 10% when adjusted for the first eight PCs (**Supplementary Table 4** and **Supplementary Figure 3**).

Mean PRS estimates by country calculated using an Empirical Bayes approach

The empirical Bayes estimates by country for the mean PRS in controls of the BCAC dataset are given in **Table 1** and **Supplementary Table 6**. Compared with the unadjusted estimates, the estimates shrunk towards the overall mean, with the shrinkage being greatest for countries that had small available sample sizes, such as Republic of North Macedonia and Russia (**Table 1**). The adjusted mean PRS by country were generally similar to those predicted by the model adjusting for six PCs (**Supplementary Table 6**). When PRSs were adjusted for the first 6 PCs, applying the empirical Bayes approach makes little difference to the estimates (**Supplementary Table 6**).

Implications for Breast Cancer Risk Prediction

To explore the effect of these differences in PRS distribution between different European populations on risk stratification, we first defined risks thresholds based on the distribution of the controls in the full BCAC dataset (**Supplementary Table 7A**). We then calculated the percentage of controls by country that would be categorized in the 90-95th, 95-99th, and >99th percentile categories, based on the distribution in the full dataset, and compared these to the percentages based on the country-specific distributions (**Supplementary Tables 7B-D**). Based on the overall distribution, approximately 4.1%, 3.7%, 1.3% and 0.5% additional women from Belarus, Republic of North Macedonia, Greece, and Italy, respectively, would be incorrectly classified in the 95-99th percentile instead at the 90-95th percentile; while 1.1% and 1.4% additional women from France and Ireland, respectively, would be incorrectly classified in the 90-95th instead of the 95-99th percentile (**Supplementary Table 7C**). **Figure 4** and **Supplementary Table 8** illustrate the PRS₃₁₃ percentile distribution in the full dataset, Greece, Italy (countries with the highest PRS₃₁₃ and including more than 100 controls) and Ireland (lowest PRS₃₁₃).

We next considered as an example a 50-year-old female from Greece with a raw PRS₃₁₃ of 0.3414 (falling into the 90th – 95th percentile category in the full BCAC dataset) and no data on family history or other known risk factors. As Greek incidence rates are not available and not currently implemented in CanRisk, we used the UK incidence rates for the calculations. If this PRS was standardized based on the mean and SD used in the CanRisk tool (when a variant call format (vcf) file is uploaded to the CanRisk tool, a raw PRS₃₁₃ can be calculated and standardized using the mean: -0.424; SD: 0.611), the individual would (assuming UK incidence rates) be given an estimate of 14.1% risk of developing breast cancer by the age of 80 and classified in the moderate risk category (**Table 2**). On the other hand, if the PRS were standardized based on the mean and SD of the controls

of Greece (mean: -0.305; SD: 0.612-raw values), she would fall in the 80-90% percentile category with an estimated 13.3% risk of developing breast cancer by the age of 80, and be classified into the population risk category (**Figure 5** and **Table 2**). Similarly, if the PRS were standardized based on the mean and SD of PRS for Greece predicted by adjustment for the first 6 PCs (mean: -0.42, SD: 0.696), she would also be classified in the population risk category (**Table 2**). Finally, if the PRS were standardized based on the mean and SD of the empirical Bayes approach (mean: -0.325 SD: 0.554), will have an estimated 13.9% risk of developing breast cancer by the age of 80, and be classified into the moderate risk category (**Table 2**).

A second example is illustrated in Table 2 and Figure 6, based on a 50-year-old female from Ireland with raw PRS₃₁₃ equal to 0.273 (equivalent to the 85th – 90th percentile-in the full BCAC dataset) and no other risk factors known. Using the CanRisk tool and assuming UK incidence rates, she would be classified in the 87.3% percentile with an estimated 13.7% absolute risk of developing breast cancer by the age of 80, which according to the NICE guidelines would be classified in the population risk category. If the PRS was standardized based on the mean and SD of PRS₃₁₃ as derived from the controls in Ireland (mean for Ireland: -0.519, and SD: 0.624 -raw values), then she would be classified in the 89.8% percentile with estimated 14.2% risk of developing breast cancer by the age of 80, classified in the moderate risk category (**Figure 6**). If the PRS were standardized based on the mean and SD of PRS for Ireland predicted by adjustment for the first 6 PCs (mean: -0.456, SD: 0.74), she would also be classified in the population risk category (**Table 2**). Finally, if the PRS were standardized based on the mean and SD of the empirical Bayes approach (mean: -0.503, SD: 0.562), will have an estimated 14.7% risk of developing breast cancer by the age of 80, and be classified into the moderate risk category (**Table 2**).

Discussion

Transferability of PRSs across different populations remains a major challenge in the field of personalized cancer risk prediction (31, 51). In this study, we explored the distribution of PRS₃₁₃ for breast cancer in European ancestry women from 21 countries, using data from studies participating in the BCAC, and further investigated how the observed variability might be accounted for in breast cancer risk prediction.

The results indicated that the PRS₃₁₃ distribution varies markedly even within Europe, with a higher mean in south-east Europe (e.g. Republic of North Macedonia, Greece, Italy) and a lower mean in western Europe (e.g. Ireland). We observed a very similar pattern in females participating in the UK Biobank, based on country of birth. If not accounted for, these differences would lead to an over- or under-estimation of risk, thus affecting the risk categorization and possibly the clinical management of some women. This may be important not only at the individual country level but also for individuals living in a different country to that of their origin.

The variability in the mean PRS₃₁₃ could not be explained by removing variants with the most variable frequency, indicating that a large number of variants may contribute to this difference. Removing such variants

to reduce the heterogeneity would not in any case be desirable as it would reduce the risk discrimination provided by the PRS. The results do, however, indicate that most if not all of the variability in the mean PRS₃₁₃ across countries in controls can be explained by adjusting for the leading ancestry informative PCs (6 PCs in the BCAC datasets, based on the OncoArray or iCOGS arrays, 8 PCs in UK Biobank). An advantage of using PCs is that they do not require any prior data from the population in question. A disadvantage, however, is that PCs require array genotyping data to generate, making them less attractive when implemented using sequencing panels. Moreover, the PCs generated using different genotyping arrays are not necessarily comparable. One interesting observation is that the heterogeneity of the ER-negative specific PRS was not eliminated even with the adjustment for 10 PCs.

We also explored generating country-specific mean PRS using an empirical Bayes approach. This approach considers both the uncertainty due to the small available sample size and the true variation in the means across the countries; these country-specific mean PRS were similar to those generated by adjusting for PCs. These values can then be used to standardise the PRS before, for example, implementing in the CanRisk tool. The risk categorization of the females from Greece and Ireland, the two examples in the Result section, was changed depending on the mean and SD of the sample used for the standardization of PRS. According to the NICE, women classified in the moderate risk category (lifetime risk of at least 17% and less than 30%), have different managing guidelines compared to women classified in the population risk category (52).

While adjustment of the PRS distribution at the population level is clearly necessary, the results raise the question as to whether it is appropriate in general to adjust PRS for PCs at the individual level, which gives different scores and potentially different risk classifications. This is a difficult question to address and hinges on whether the PCs should be regarded as nuisance parameters correcting for confounding factors, such as screening or lifestyle factors. Reanalysis of prospective studies with BCAC OncoArray dataset shows that the first two PCs are associated with the PRS (PC1 negatively, PC2 positively) and are also associated with risk (in the same direction). The PRS effect size (OR per 1 SD) is essentially unchanged whether or not adjustment for PCs is made (Supplementary Table 9 and Supplementary Table 10). This implies that risk discrimination would be slightly improved by including the effect of PCs in the PRS, and that adjusting the PRS for PCs further reduces the discrimination. Fortunately, the association between the PC1 and risk is weak and, within a country, the variation in the PC1 is not large enough to materially change risk categories.

The differences in the PRS distribution across Europe are a manifestation, on a continental scale, of the larger intercontinental differences – the mean PRS is higher in both east Asian and African populations than in the European dataset examined here (28, 29, 53). It is interesting to note that the pattern appears unrelated to the population-specific incidence, which is fact lower in south-east than north-west Europe (54), presumably because the effect on disease incidence is counterbalanced by larger effects of lifestyle (or other genetic) factors. It remains unclear whether the differences in the PRS can be attributed purely to random genetic drift or whether selection pressures relevant to breast cancer aetiology are involved.

We would like to acknowledge some potential limitations of our study. The dataset we used was genetically-homogeneous and maybe not completely representative of the population of each country. It remains an important issue how to interpret the PRS in individuals classified as mixed ancestry. In the future, the exploration of the distribution of the mean PRS across the individuals classified as mixed ancestry could be performed. Furthermore, evaluation of the country-specific calibrated PRS in combination with classical breast cancer risk factors should be performed in order to explore the extend to these findings have on final risk prediction.

In summary, these results demonstrate that the implementation of the PRS313 in risk prediction models such as CanRisk/BOADICEA could potentially require country-specific calibration. This can be achieved by genotyping a large control group to obtain population-specific means, by using a principal components adjustment, or the empirical Bayes approach described here.

Figures

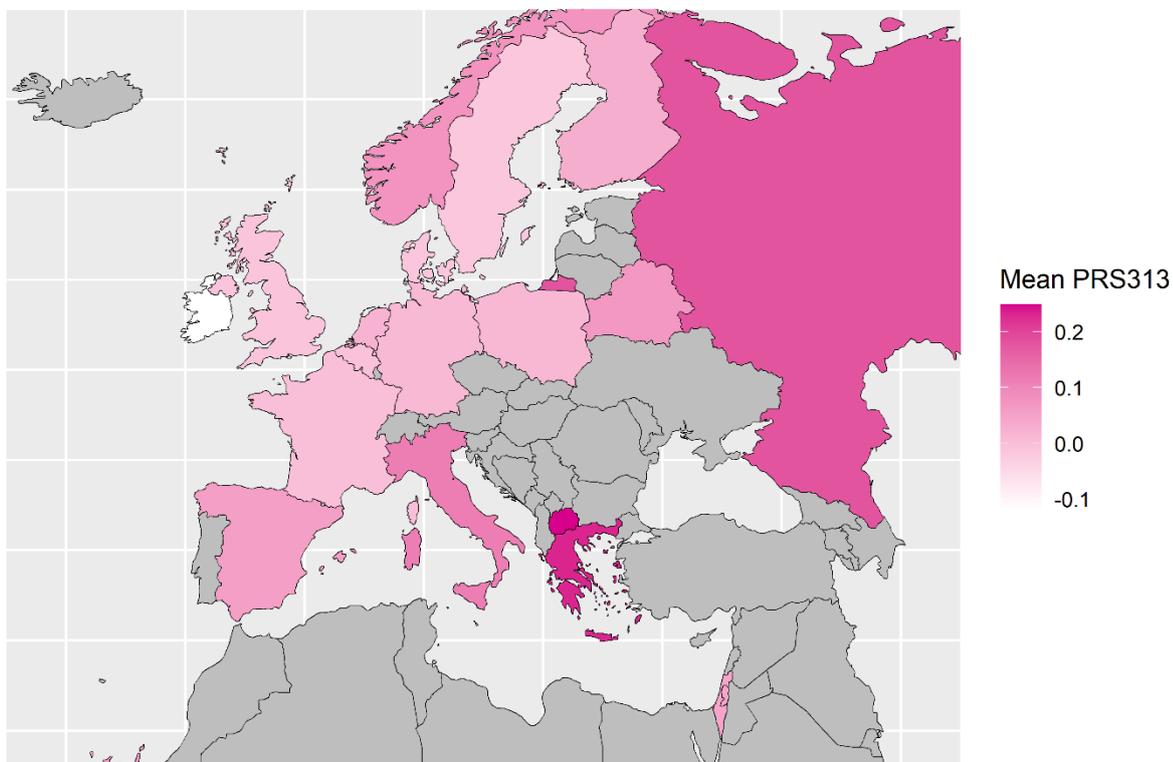


Figure 1: Map of the European countries of origin of BCAC study participants included in the analysis. Countries were coloured based on their mean standardized PRS₃₁₃ in control dataset of BCAC. Countries with higher mean are represented with darker colour while those with lower mean with lighter colour.

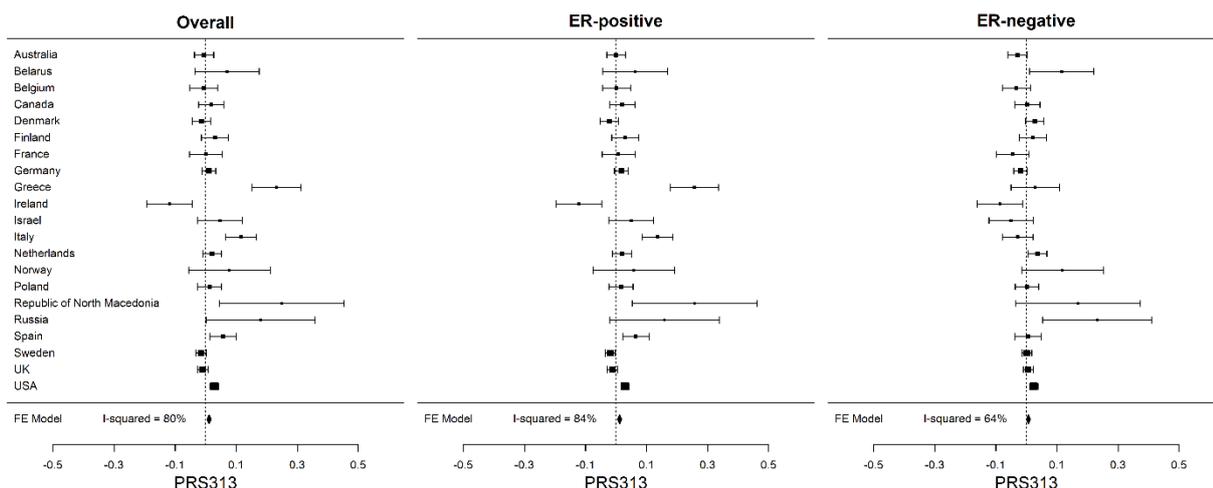


Figure 2: Distribution of the standardized PRS₃₁₃ across country of origin for overall, ER-positive and ER-negative breast cancer in control dataset of BCAC. The squares represent the mean PRS by country and the error bars represent the corresponding 95% confidence intervals (FE Model: Fixed effect Model).

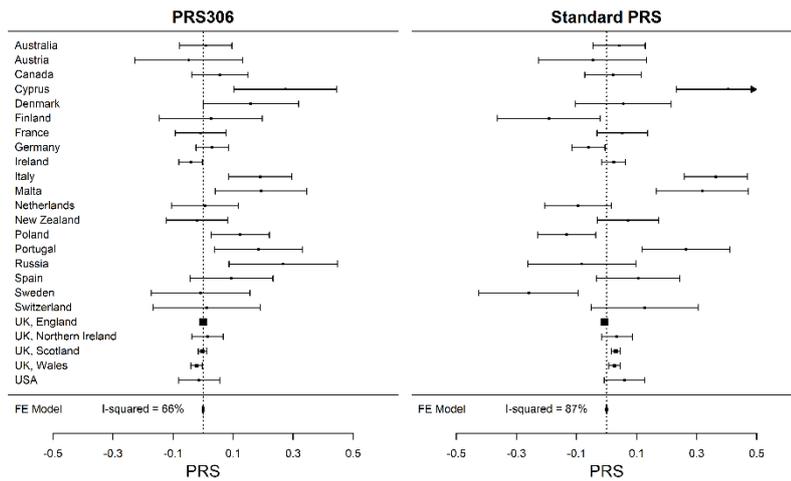


Figure 3: Distribution of the mean PRS₃₀₆, and “standard” PRS for breast cancer, as defined in the UK Biobank, across countries of origin of participating white females. The squares represent the mean PRS by country and the error bars represent the corresponding 95% confidence intervals (FE Model: Fixed effect Model).

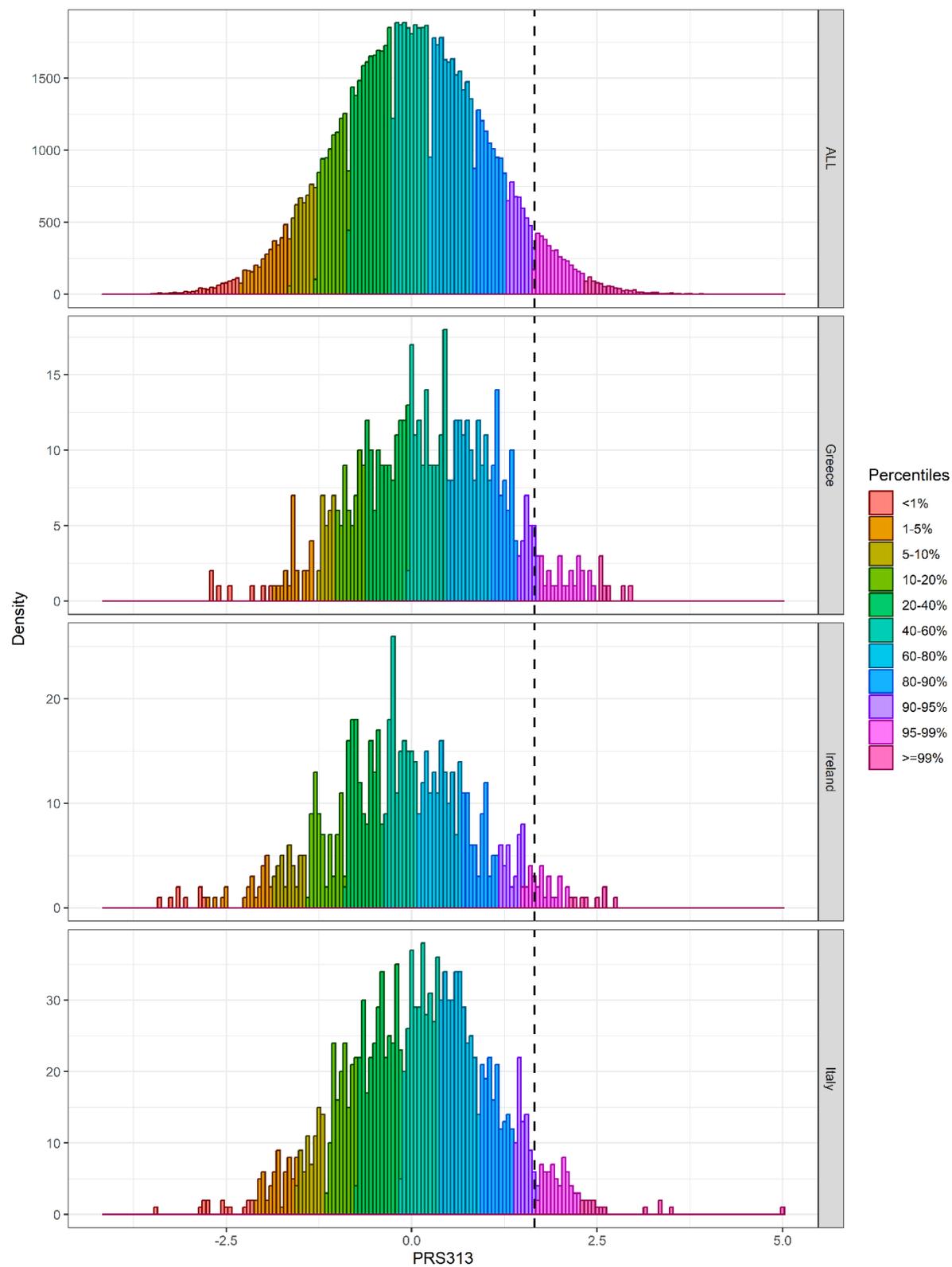
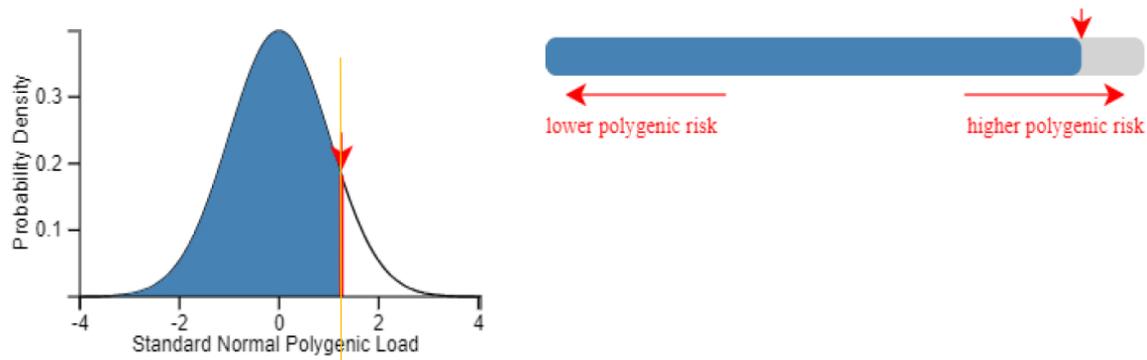


Figure 4: PRS₃₁₃ distribution in controls by percentiles in the pooled BCAC dataset, Greece, Ireland and Italy. The dashed line corresponds to the 95th percentile of the PRS₃₁₃ distribution in controls of the pooled BCAC dataset.

Breast Cancer PRS: 89.5 of people in the population have a **lower** polygenic load. 10.5 of people in the population have a **higher** polygenic load. undefined alpha:0.441 z-score:1.2527



Breast Cancer PRS: 85.5 of people in the population have a **lower** polygenic load. 14.5 of people in the population have a **higher** polygenic load. undefined alpha:0.441 z-score:1.056013

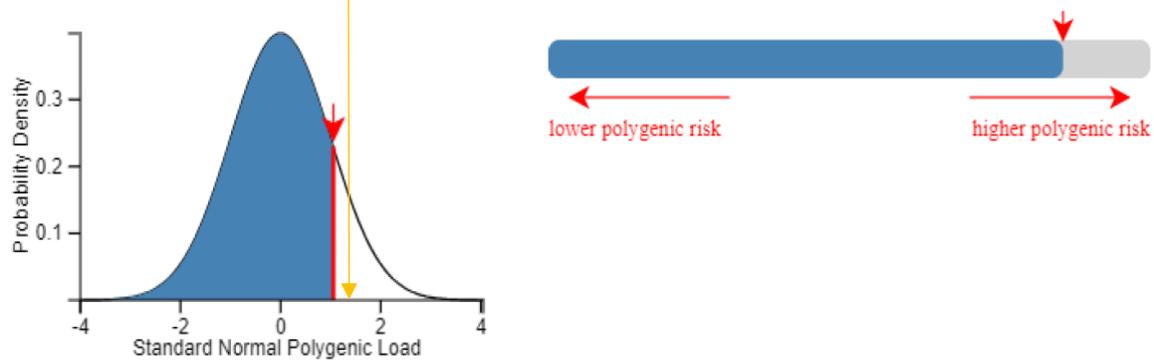
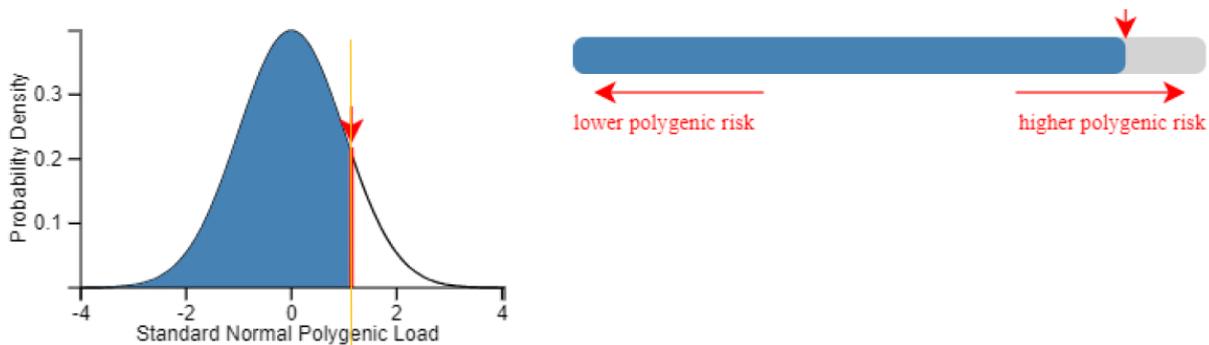


Figure 5: Classification of a 50-year-old woman from Greece when her raw PRS₃₁₃, which is equal to 0.34 is standardized based on the mean and SD of the controls of BOADICEA model (upper panel) and Greece (lower panel), using the CanRisk tool. Plots were generated using the CanRisk tool (www.canrisk.org).

Breast Cancer PRS: 87.3 of people in the population have a **lower** polygenic load. 12.7 of people in the population have a **higher** polygenic load. undefined alpha:0.441 z-score:1.141071



Breast Cancer PRS: 89.8 of people in the population have a **lower** polygenic load. 10.2 of people in the population have a **higher** polygenic load. undefined alpha:0.441 z-score:1.270339

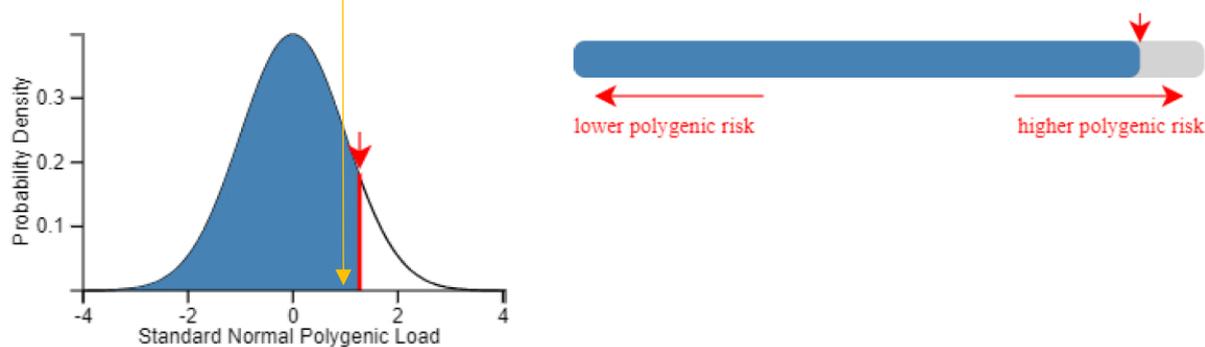


Figure 6: Classification of a 50-year-old woman from Ireland when her raw PRS₃₁₃, which is equal to 0.27 is standardized based on the mean and SD of the controls BOADICEA (upper panel) model and Ireland (lower panel), using the CanRisk tool (www.canrisk.org).

Tables

Table 1: Mean standardized PRS₃₁₃ by country in controls in the pooled BCAC dataset, estimated when adjusted for array, 6 PCs country and array, using fitted values adjusted for 6 PCs and array and when using an Empirical Bayes approach adjusted for array.

Country	Number of Controls	Mean PRS ₃₁₃ ¹	Mean PRS adjusted for array and 6 PCs	PRS adjusted for 6 PCs, fitted values ²	Empirical Bayes Posterior Mean ³
Australia	4049	-0.005	0.01	-0.005	-0.003
Belarus	342	0.07	0.071	0.016	0.064
Belgium	1823	-0.006	-0.007	0.010	-0.002
Canada	2277	0.018	0.019	0.013	0.02
Denmark	5241	-0.013	0.012	-0.031	-0.012
Finland	2083	0.031	0.008	0.010	0.032
France	1372	0.0003	-0.008	0.008	0.004
Germany	8563	0.011	0.004	0.013	0.011
Greece	607	0.232	0.043	0.208	0.199
Ireland	719	-0.118	-0.015	-0.112	-0.092
Israel	724	0.047	0.001	0.062	0.047
Italy	1554	0.115	-0.007	0.131	0.11
Netherlands	4407	0.021	0.043	-0.019	0.022
Norway	217	0.077	0.094	-0.027	0.066
Poland	2554	0.013	0.025	0.010	0.015
Republic of North Macedonia	92	0.25	0.134	0.140	0.129
Russia	120	0.18	0.166	0.044	0.11
Spain	2098	0.057	-0.006	0.057	0.056
Sweden	16680	-0.015	0.005	-0.017	-0.014
UK	16854	-0.01	0.019	-0.023	-0.01
USA	21696	0.029	0.033	0.013	0.029

¹Mean PRS₃₁₃ adjusted for array

²Mean PRS₃₁₃ by country using predicted PRS of each individual; estimated using linear predictor of PRS vs 6 PCs and the command predict () in R.

³Country-specific estimates, means β , using the Empirical Bayes approach, adjusted for array

Table 2: Mean and SD used to standardize PRS₃₁₃ of a 50-year-old woman with raw PRS₃₁₃ equal to 0.341 from Greece and another 50-year-old woman with raw PRS₃₁₃ equal 0.273 from Ireland, and the risk estimation and categorization when using the CanRisk tool, Greek and Ireland values.

Samples used for the standardization:	Raw PRS; Mean (SD)	Standardized PRS ¹	Percentage based on CanRisk tool	Lifetime risk based on CanRisk tool ²	NICE category	Risk
Individual from Greece with raw PRS₃₁₃ = 0.341 (falling into the 90–95% percentile category in the full BCAC dataset)						
CanRisk tool ³	-0.424 (0.611)	1.253	89.5%	14.1%	Moderate	
Controls Greece (raw) ⁴	-0.305 (0.612)	1.056	85.5%	13.3%	Population	
Controls Greece adjusted for 6 PCs (raw)	-0.420 (0.696)	1.094	86.3%	13.5%	Population	
Controls Greece, using Empirical Bayes method	-0.325 (0.554)	1.204	88.6%	13.9%	Moderate	
Individual from Ireland with raw PRS₃₁₃ = 0.273 (falling into the 85–90% percentile category in the full BCAC dataset)						
CanRisk tool ³	-0.424 (0.611)	1.14	87.3%	13.7%	Population	
Controls Ireland (raw) ¹	-0.519 (0.624)	1.27	89.8%	14.2%	Moderate	
Controls Ireland adjusted for 6 PCs (raw)	-0.456 (0.74)	0.985	83.8%	13%	Population	
Controls Ireland, using EB	-0.503 (0.562)	1.38	91.7%	14.7%	Moderate	

¹Standardised based on the mean and SD specified in the second column

²Absolute risk of developing breast cancer by the age of 80

³When a variant call format (vcf) file is uploaded to the CanRisk tool, a raw PRS₃₁₃ can be calculated and standardized using the mean (SD) -0.424 (0.611)

⁴Adjusted for array type

Author contributions: Writing Group: K.Y., K.Mi., D.F.E., A.C.A., N.Ma, and J. Si.; Study design: K.Mi., D.F.E., A.C.A., N.Ma., J.Si., and K.Y.; Data management: M.K.B., Q.W.; Statistical Analysis: K.Y., N.Ma, J.D., M.Z., D.F.E., K.Mi.; Provided data: M.A., T.U.A., I.L.A., H.A-C., N.N.A., V.A., K.J.A., A.Au., A.Bat., S.Be., M.Berm., A.Ber., K.Bia., N.B., C.Bo., N.V.B., S.E.B., K.Br., H.Bra., H.Bre., N.J.C., F.C., J.E.C., J.C-C., G.C-T., W.K.C., NBCS Collaborators, S.V.C., F.J.C., A.Cox., S.S.C., K.Cz., M.B.D., P.D., T.D., A.M.D., D.M.E., A.H.E., C.En., M.E., D.G.E., P.A.F., O.F., H.F., M.G-D., A.G-M., A.G-N., P.Gu., E.Hah., C.A.H., P.Hall., U.H., J.M.H., V.H., J.H., A.Hol., E.Hon., M.J.H., R.H., J.L.Ho., S.H., A.How., ABCTB Investigators, kConFab Investigators, S.J., A.Jak., H.J., N.J., R.Ka., E.K.K., C.M.Ki., S.Kou., V.N.K., J.V.L., D.La., F.Lej., A.Lin., M.Lus., R.J.M., A.Man., D.M., U.M., R.L.M., R.A.M., H.Ne., N.Ob., K.Of., T-W.P-S., A.V.P., C.P., P.Pe., P.D.P.P., G.Pi., D.P.K., K.Py., P.Ra., M.U.R., G.R., E.R., J.R, A.Ro., E.H.R., E.S., D.P.S., E.J.S., M.K.S., R.K.S., C.Sc., X-O.S., M.C.S., J.St., J.A.T., L.R.T., C.M.V, I.VDB., W.W., R.Wi., W.Z., J.Si., A.C.A, D.F.E.. All authors read and approved the final version of the manuscript.

References

1. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature genetics*. 2013;45(4):353-61.
2. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nature genetics*. 2015;47(4):373-80.
3. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*. 2017;551(7678):92-4.
4. Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nature genetics*. 2020;52(6):572-81.
5. Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C, Wahlström C, et al. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *The New England journal of medicine*. 2021;384(5):428-39.
6. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Blom MJ, et al. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *JAMA - Journal of the American Medical Association*. 2017;317(23):2402-16.
7. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nature protocols*. 2020;15(9):2759-72.
8. Wand H, Lambert SA, Tamburro C, Iacocca MA, O'Sullivan JW, Sillari C, et al. Improving reporting standards for polygenic scores in risk prediction studies. *Nature: Nature Research*; 2021. p. 211-9.
9. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *Journal of the National Cancer Institute*. 2015;107(5).
10. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics*. 2018;50(9):1219-24.
11. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *American journal of human genetics*. 2019;104(1):21-34.
12. Shieh Y, Eklund M, Madlensky L, Sawyer SD, Thompson CK, Stover Fiscalini A, et al. Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. *Journal of the National Cancer Institute*. 2017;109(5).
13. Pashayan N, Morris S, Gilbert FJ, Pharoah PDP. Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. *JAMA Oncol*. 2018;4(11):1504-10.
14. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genetics in medicine : official journal of the American College of Medical Genetics*. 2019;21(8):1708-18.

15. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome medicine*. 2020;12(1):44.
16. Pashayan N, Antoniou AC, Ivanus U, Esserman LJ, Easton DF, French D, et al. Personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nature reviews Clinical oncology*. 2020;17(11):687-705.
17. Brooks JD, Nabi HH, Andrulis IL, Antoniou AC, Chiquette J, Després P, et al. Personalized Risk Assessment for Prevention and Early Detection of Breast Cancer: Integration and Implementation (PERSPECTIVE I&D). *Journal of personalized medicine*. 2021;11(6).
18. van den Broek JJ, Schechter CB, van Ravesteyn NT, Janssens A, Wolfson MC, Trentham-Dietz A, et al. Personalizing Breast Cancer Screening Based on Polygenic Risk and Family History. *Journal of the National Cancer Institute*. 2021;113(4):434-42.
19. Pashayan N, Easton DF, Michailidou K. Polygenic risk scores in cancer screening: a glass half full or half empty? *The Lancet Oncology*. 2023;24(6):579-81.
20. Yang X, Kar S, Antoniou AC, Pharoah PDP. Polygenic scores in cancer. *Nature reviews Cancer*. 2023;23(9):619-30.
21. Carver T, Hartley S, Lee A, Cunningham AP, Archer S, Babb de Villiers C, et al. CanRisk Tool-A Web Interface for the Prediction of Breast and Ovarian Cancer Risk and the Likelihood of Carrying Genetic Pathogenic Variants. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2021;30(3):469-73.
22. Archer S, Babb de Villiers C, Scheibl F, Carver T, Hartley S, Lee A, et al. Evaluating clinician acceptability of the prototype CanRisk tool for predicting risk of breast and ovarian cancer: A multi-methods study. *PLoS One*. 2020;15(3):e0229999.
23. Lakeman IMM, Rodríguez-Girondo M, Lee A, Rutter R, Stricker BH, Wijnant SRA, et al. Validation of the BOADICEA model and a 313-variant polygenic risk score for breast cancer risk prediction in a Dutch prospective cohort. *Genetics in Medicine*. 2020.
24. Pal Choudhury P, Brook MN, Hurson AN, Lee A, Mulder CV, Coulson P, et al. Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort of women of European ancestry. *Breast cancer research : BCR*. 2021;23(1):22.
25. Li SX, Milne RL, Nguyen-Dumont T, Wang X, English DR, Giles GG, et al. Prospective Evaluation of the Addition of Polygenic Risk Scores to Breast Cancer Risk Models. *JNCI cancer spectrum*. 2021;5(3).
26. Yang X, Eriksson M, Czene K, Lee A, Leslie G, Lush M, et al. Prospective validation of the BOADICEA multifactorial breast cancer risk prediction model in a large prospective cohort study. *Journal of medical genetics*. 2022;59(12):1196-205.
27. Lee A, Mavaddat N, Cunningham A, Carver T, Ficorella L, Archer S, et al. Enhancing the BOADICEA cancer risk prediction model to incorporate new data on RAD51C, RAD51D, BARD1 updates to tumour pathology and cancer incidence. *Journal of medical genetics*. 2022;59(12):1206-18.
28. Ho WK, Tan MM, Mavaddat N, Tai MC, Mariapun S, Li J, et al. European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. *Nature communications*. 2020;11(1):3833.
29. Du Z, Gao G, Adedokun B, Ahearn T, Lunetta KL, Zirpoli G, et al. Evaluating Polygenic Risk Scores for Breast Cancer in Women of African Ancestry. *Journal of the National Cancer Institute*. 2021;113(9):1168-76.
30. Liu C, Zeinomar N, Chung WK, Kiryluk K, Gharavi AG, Hripcsak G, et al. Generalizability of Polygenic Risk Scores for Breast Cancer Among Women With European, African, and Latinx Ancestry. *JAMA network open*. 2021;4(8):e2119084.
31. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature genetics*. 2019;51(4):584-91.
32. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *American journal of human genetics*. 2017;100(4):635-49.
33. Ding Y, Hou K, Xu Z, Pimplaskar A, Petter E, Boulier K, et al. Polygenic scoring accuracy varies across the genetic ancestry continuum. *Nature*. 2023;618(7966):774-81.

34. Kachuri L, Chatterjee N, Hirbo J, Schaid DJ, Martin I, Kullo IJ, et al. Principles and methods for transferring polygenic risk scores across global populations. *Nature reviews Genetics*. 2023.
35. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2017;26(1):126-35.
36. O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, et al. A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS genetics*. 2014;10(4):e1004234.
37. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
38. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-9.
39. 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;12(3):e1001779.
40. Li Y, Byun J, Cai G, Xiao X, Han Y, Cornelis O, et al. FastPop: a rapid principal component derived method to infer intercontinental ancestry using genetic data. *BMC Bioinformatics*. 2016;17:122.
41. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics*. 2016;48(10):1279-83.
42. Thompson DJ, Wells D, Selzam S, Peneva I, Moore R, Sharp K, et al. UK Biobank release and systematic evaluation of optimised polygenic risk scores for 53 diseases and quantitative traits. *medRxiv*. 2022:2022.06.16.22276246.
43. Schmidt MK, Hogervorst F, van Hien R, Cornelissen S, Broeks A, Adank MA, et al. Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2016;34(23):2750-60.
44. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*. 1987;43(3):671-81.
45. Team RC. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2023.
46. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*. 2010;36(3):1 - 48.
47. South A. Rnaturalearth: world map data from natural earth. R package version 01 0. 2017.
48. South A. rnaturalearthdata: world vector map data from Natural Earth used in'rnaturalearth'. R package version 0.1. 0. 2017.
49. Pebesma EJ. Simple features for R: standardized support for spatial vector data. *R J*. 2018;10(1):439.
50. Bivand R, Rundel C, Pebesma E, Stuetz R, Hufthammer KO, Bivand MR. Package 'rgeos'. The Comprehensive R Archive Network (CRAN). 2017.
51. Wang Y, Tsuo K, Kanai M, Neale BM, Martin AR. Challenges and Opportunities for Developing More Generalizable Polygenic Risk Scores. *Annual review of biomedical data science*. 2022;5:293-320.
52. National Institute for Health and Care Excellence: Guidelines. Familial breast cancer: classification, care and managing breast cancer and related risks in people with a family history of breast cancer. London: National Institute for Health and Care Excellence (NICE)

Copyright © NICE 2020.; 2019.

53. Ho WK, Tai MC, Dennis J, Shu X, Li J, Ho PJ, et al. Polygenic risk scores for prediction of breast cancer risk in Asian populations. *Genetics in medicine : official journal of the American College of Medical Genetics*. 2022;24(3):586-600.
54. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2021.