Supplementary Methods

We developed a simulation-based risk prediction framework for estimating the distribution of invasive breast cancer risk in the general United Kingdom (UK) population. Specifically, risk profiles of 1,000,000 women were generated using probability distributions calculated based on the distribution of risk factors for women at two different ages (i.e. 40 and 50 years old), and published relative risks (RRs), adjusted by other risk factors when available. Risk factors included age at menarche, parity, age at first live birth, current OC use, current combined menopausal replacement therapy (MHT) use, benign breast disease (BBD), body mass index (BMI), alcohol intake, smoking, family history in first degree relatives, mammographic breast density, and polygenic risk score (PRS).

The distribution of risk factors within each age group was generated according to a decomposition of the joint probability distribution by sets of risk-factors that are most correlated among themselves. Specifically, the joint probabilies for women aged 40 and 50 were decomposed as follows:

Pr ([PRS, family history, BMI, breast density, age at menarche, alcohol intake, BBD, OC use, parity, age at first live birth, density]|age=40) = Pr(PRS, family history) * Pr(BMI) * Pr(breast density|BMI) * Pr(age at menarche) * Pr(OC) * Pr(BBD) * Pr(parity) * Pr(age at first birth| parity) * Pr(alcohol intake, smoking)

and

Pr ([PRS, family history, BMI, breast density, age at menarche, alcohol intake, BBD, HRT use, parity, age at first live birth]|age=50) = Pr(PRS, family history) * Pr(BMI) * Pr(breast density|BMI) * Pr(age at menarche) * Pr(HRT) * Pr(BBD) * Pr(parity) * Pr(age at first birth| parity) * Pr(alcohol intake, smoking)

Supplementary Table 1 shows the sources of information for different risk factors. When possible, we used population-based data on risk factor distributions representative of the UK population. This included OC use, BMI, alcohol intake, and smoking from the Health Survey for England[1,2] and parity and age at first birth from Cohort Fertility Tables in England and Wales from the Office of National Statistics (ONS). Distributions for other risk factors (i.e. density, family history and BBD) were derived from published reports. Correlation between

risk factors were investigated in sources from which multiple risk factors were obtained: pairwise correlations between OC use, BMI, alcohol intake, and smoking were checked within the Health Survey for England, and only alcohol intake and smoking had a statistically significant (P>0.05) correlation. The correlation between BMI and breast density was calculated using the reported mean and SD of BMI by density category from the NCI Breast Cancer Surveillance Consortium[3] BMI categories within each density category were sampled assuming a normal distribution, and the distribution of breast density conditional on BMI category was derived. Age at first live birth was simulated within parous women; however correlations between number of births and age at first live birth could not be ascertained from cohort fertility tables for England and Wales and was thus not included in the model.

The joint distribution of PRS and family history in the population and the joint RR for breast cancer associated with these factors were obtained based on an analytic formula[4] that assumes a log-normal distribution for PRS[5] conditional on family history where the mean of the log-normal distribution is allowed to vary by family history and the standard deviation (SD) is fixed. We used two theoretical distributions of the PRS in our calculations: one for a 76-SNP PRS with SD=0.46 (calculated assuming that the 76-SNP PRS explains 15% of the familial risk[6]); and the other for an improved PRS with SD=0.65 (calculated assuming a PRS that explains 30% of the familial risk[6]). The SDs were calculated as the square root of the product of a variance of 1.44 (SD=1.2[7]) times the percentage of familial risk (sibling RR=2.0) explained by the PRS. We chose these two PRS distributions to illustrate a range of possible scenarios, one that is attainable today based on 76 established susceptibility loci, and one that could be attained in the near future after the completion of ongoing, large-scale genotyping projects. For all other risk factors, RR estimates by age were obtained preferably from meta-analyses and adjusted for other risk factors included in this model. Most RR estimates for women aged 40 years were obtained from a single source.[8]

The joint RR of PRS, family history and all other risk factors were calculated assuming multiplicative effects across these factors. The means of the RR distributions across models were rescaled so that the simulated population mean (with respect to the average risk of the population) was equal to 1. This rescaled RR was then used as a multiplier for age-specific incidence for the calculation of absolute risk of breast cancer. A

person's RR was assumed to be fixed overtime. The lifetime risk from age 20-80 years, and 10-year absolute risk of invasive breast cancer for 50-year old women were calculated using the Gail calculation method with averaged incidence and mortality rates in the UK for 2006-2010 (Office for National Statistics, ONS[9]), and calibrated to the average absolute risk in the population (RR=1).

To plot the percent of cases captured by different percent of population at highest risk, we stratified the simulated population into 100 groups according to RR, in 0.1 increments, The RR ranges (min-max) corresponding to the eight different models in women aged 50 years old were:

Model 1 (Qx risk factors): RR of 0.36 - 13.7

Model 2 (Qx risk factors + density): RR of 0.15 - 15.0

Model 3 (76-SNP PRS alone): RR of 0.28 - 3.20

Model 4 (Qx risk factors + 76-SNP PRS): RR of 0.10 - 27.4

Model 5 (Qx risk factors + 76-SNP PRS + density): RR of 0.05 - 26.5

Model 6 (Improved PRS alone): RR of 0.18 - 4.80

Model 7 (Qx risk factors + Improved PRS): RR of 0.05 - 37.8

Model 8 (Qx risk factors + Improved PRS + density): RR of 0.02 - 35.5

The expected percentage of cases within each RR group was obtained by computing the product of the frequency of the group times the average absolute risk for that group, divided by the average population absolute risk. Thus, for a given RR group, the population at highest risk was the cumulative frequency of all RR groups at equal or higher risk, and the percent of cases was the cumulative percentage of cases explained in all RR groups at equal or higher risk.

Supplementary Table 1: Sources of information on distributions of risk factors in the UK population, and relative risks for breast cancer

	Population Distribution		Relative risk of breast cancer		Reference for Distribution	Reference for Relative Risk
Risk factor	**Age 40**	**Age 50**	**Age 40**	**Age 50**		
Age at menarche						
<11 years	3.8%	3.8%		1.19	[10]	[10]
11 years	12.3%	12.3%		1.09		
12 years	20.1%	20.1%		1.07		
13 years	27.2%	27.2%		1.00		
14 years	17.4%	17.4%	1.00	0.98		
15 years	10.3%	10.3%	0.77	0.92		
>=16 years	8.9%	8.9%		0.82		
Parity						
Nulliparous	20.0%	20.0%	1.08	1.00	[11]	[8,12]
Parous	80.0%		1.00			
1 births		14.0%	N/A	0.87		
2 births		38.0%	N/A	0.81		
3+ births		28.0%	N/A	0.71		
Age at first live birth						
<=20 years	13.0%	11.0%	0.78	1.00	[11]	[12]
<=25 years	20.0%	24.0%	0.87	1.01		
<30 years	17.0%	23.0%	1.00	1.11		
>=30 tears	30.0%	22.0%	1.02	1.24		
Current use of OC						
No	87.0%	N/A	1.00	N/A	[1]	[8]
Yes	13.0%	N/A	1.30	N/A		
Current use of estrogen and Progesterone HRT						
No	N/A	80.0%	N/A	1.00	[13]	[12]
Yes	N/A	20.0%	N/A	1.65		
BBD						
No BBD	90.0%	86.7%	1.00	1.00		
Proliferative with no atypia	9.0%	12.0%	1.51	1.50	(Assumed based on published incidence estimates[14,15])	[16,17]
Atypical hyperplasia	1.0%	1.3%	4.00	2.63		
BMI						
< 18.5	0.6%		1.28		[1]	[8,12]
18.5 -< 25	38.8%	34.3%	1.00	0.82		
25 -< 30	36.0%	34.8%	0.92	1.00		
>= 30	24.6%	30.9%	0.74	1.18		

Alcohol intake						
0 g per day	16.8%	13.2%	1.00	1.00	[1]	[18]
<5 g per day	12.0%	15.3%	1.01	1.01		
5-14 g per day	36.9%	32.4%	1.03	1.03		
15-24 g per day	15.6%	14.1%	1.13	1.13		
25-34 g per day	8.2%	8.8%	1.21	1.21		
35-44 g per day	6.4%	8.9%	1.32	1.32		
>45 g per day	4.0%	7.3%	1.46	1.46		
Smoking status						
Never	58.4%	54.3%	1.00	1.00	[1]	[18]
Former	20.3%	23.0%	1.09	1.09		
Current	21.3%	22.7%	1.12	1.12		
Family history of breast cancer in first-degree relatives						
No	0.0%	0.0%	1.00	1.00	[19]	[20]
Yes	7.2%	10.0%	2.30	1.80		
Mammographic breast density *						
BI-RADS 1	4.1%	7.4%	0.41	1.00	[21]	[8,22]
BI-RADS 2	35.2%	46.6%	1.00	2.04		
BI-RADS 3	47.0%	39.6%	1.75	2.81		
BI-RADS 4	13.7%	6.4%	2.33	4.08		

* BI-RADS classification of mammographic breast density into the following four categories:
BI-RADS 1: Almost entirely fatty (<25% fibrous and glandular tissue).
BI-RADS 2: Scattered fibroglandular densities (25%-50% fibrous and glandular tissue).
BI-RADS 3: Heterogeneously dense (51%-75% fibrous and glandular tissue).
BI-RADS 4: Extremely dense (>75% fibrous and glandular tissue.

Supplementary Table 2: Identification of women aged 40 years in a UK population at moderate and high-risk of invasive breast cancer (defined as RR >2.0-3.0 and RR>3.0, respectively, compared to the population average), for different combinations of risk factors and two polygenic risk scores (PRS). The following parameters are shown for eight risk prediction models: AUC, the % of the population found at moderate and high levels of risk according to the different models, and the % of cases in the population expected to occur among women at these levels of risk.

	Model 1: Qx risk factors*		**Model 2: Qx risk factors + density**
AUC	0.608		0.653
% Population (% cases) at different risk thresholds**			
Moderate risk (RR >2.0 - 3.0 or life-time risk >19.4%-27.5%)	3.8 (8.5)		4.7 (10.5)
High risk (RR > 3.0 or life-time risk >27.5%)	1.3 (4.8)		1.9 (7.4)
Combined (RR>2.0 or life-time risk >19.4%)	*5.1 (13.3)*		*6.6 (17.9)*
	Model 3: 76-SNP PRS	**Model 4:Qx risk factors + 76-SNP PRS**	**Model 5: Qx risk factors + density +76-SNP PRS**
AUC	0.628	0.665	0.689
% Population (% cases) at different risk thresholds**			
Moderate risk (RR >2.0 - 3.0 or life-time risk >19.4%-27.5%)	4.2 (9.1)	5.7 (12.9)	6.7 (15.3)
High risk (RR > 3.0 or life-time risk >27.5%)	1.1 (3.4)	2.4 (9.6)	3.3 (13.5)
Combined (RR>2.0 or life-time risk >19.4%)	*5.2 (12.5)*	*8.1 (22.5)*	*10.0 (28.8)*
	Model 6: Improved PRS	**Model 7: Qx risk factors + Improved PRS**	**Model 8: Qx risk factors + density + Improved PRS**
AUC	0.677	0.701	0.711
% Population (% cases) at different risk thresholds**			
Moderate risk (RR >2.0 - 3.0 or life-time risk >19.4%-27.5%)	6.3 (14.0)	6.7 (15.2)	7.0 (16.2)
High risk (RR > 3.0 or life-time risk >27.5%)	3.2 (11.7)	3.8 (16.2)	4.6 (20.2)
Combined (RR>2.0 or life-time risk >19.4%)	*9.5 (25.7)*	10.5 (31.4)	*11.7 (36.4)*

*Questionnaire (Qx) based risk factors include age at menarche, parity, age at first birth, hormone use (oral contraceptives), BMI, BBD, alcohol intake, smoking and family history of breast cancer in first degree relatives.

** Life-time risk (from age 20 to 80 years), 10-year and 5-year risk thresholds corresponding to RR of 2.0 and 3.0 for a women aged 50 years old: RR = 2.0: 19.4% life-time risk, 3.1% 10-year risk and 1.2% 5-year risk; RR = 3.0: 27.5% life-time risk, 4.6% 10-year risk and 1.8% 5-year risk.

References

1. *Health Survey for Enlgand*. Available at: http://data.gov.uk/dataset/health_survey_for_england. Accessed July 10, 2013.

2. Mindell J, Biddulph JP, Hirani V, et al. Cohort profile: the health survey for England. *International Journal of Epidemiology*. 2012;41(6):1585–1593. doi:10.1093/ije/dyr199.

3. Vacek PM, Geller BM. A prospective study of breast cancer risk using routine mammographic breast density measurements. *Cancer Epidemiol Biomarkers Prev*. 2004;13(5):715–722.

4. Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park J-H. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat Genet*. 2013;45(4):400–5– 405e1–3. doi:10.1038/ng.2579.

5. Pharoah PDP, Antoniou AC, Easton DF, Ponder BAJ. Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med*. 2008;358(26):2796–2803. doi:10.1056/NEJMsa0708739.

6. *COGS Primer*. Available at: http://www.nature.com/icogs/primer/common-variation-and-heritability-estimates-for-breast-ovarian-and-prostate-cancers/. Accessed May 30, 2013.

7. Pharoah PDP, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BAJ. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet*. 2002;31(1):33–36. doi:10.1038/ng853.

8. Nelson HD, MD, MPH, et al. Risk Factors for Breast Cancer for Women Age 40 to 49: A Systematic Review and Meta-analysis. *Ann Intern Med*. 2012;9(156):635–648. doi:10.1059/0003-4819-156-9-201205010-00006.

9. *Office for National Statistics*. ONS Available at: http://www.ons.gov.uk/ons/index.html. Accessed June 19, 2013.

10. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *The Lancet Oncology*. 2012;13(11):1141–1151. doi:10.1016/S1470-2045(12)70425-4.

11. *Cohort Fertility Tables in England and Wales*. Office for National Statistics Available at: http://data.gov.uk/dataset/cohort_fertility_england_and_wales. Accessed June 19, 2013.

12. Reeves GK, Pirie K, Green J, Bull D, Beral V, Million Women Study Collaborators. Comparison of the effects of genetic and environmental risk factors on in situ and invasive ductal breast cancer. *Int J Cancer*. 2012;131(4):930–937. doi:10.1002/ijc.26460.

13. Parkin DM. Is the recent fall in incidence of post-menopausal breast cancer in UK related to changes in use of hormone replacement therapy? *Eur J Cancer*. 2009;45(9):1649–1653. doi:10.1016/j.ejca.2009.01.016.

14. Rohan TE, Miller AB. A cohort study of oral contraceptive use and risk of benign breast disease. *Int J Cancer*. 1999;82(2):191–196.

15. Goldacre MJ, Abisgold JD, Yeates DGR, Vessey MP. Benign breast disease and subsequent breast cancer: English record linkage studies. *J Public Health (Oxf)*. 2010;32(4):565–571. doi:10.1093/pubmed/fdq001.

16. Hartmann LC, Sellers TA, Frost MH, et al. Benign breast disease and the risk of breast cancer. *N Engl J Med*. 2005;353(3):229–237. doi:10.1056/NEJMoa044383.

17. Byrne C, Schairer C, Brinton LA, et al. Effects of mammographic density and benign breast disease on breast cancer risk (United States). *Cancer Causes Control*. 2001;12(2):103–110.

18. Alcohol, tobacco and breast cancer – collaborative reanalysis of individual data from 53 epidemiological studies, including 58515 women with breast cancer and 95067 women without the disease. *Br J Cancer*. 2002;87(11):1234–1245. doi:10.1038/sj.bjc.6600596.

19. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet*. 2001;358(9291):1389–1399. doi:10.1016/S0140-6736(01)06524-2.

20. Pharoah PD, Day NE, Duffy S, Easton DF, Ponder BA. Family history and the risk of breast cancer: a systematic review and meta-analysis. *Int J Cancer*. 1997;71(5):800–809.

21. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med*. 2008;148(5):337–347.

22. McCormack VA, Santos Silva Dos I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev*. 2006;15(6):1159–1169. doi:10.1158/1055-9965.EPI-06-0034.