ONLINE SUPPLEMENTARY MATERIAL

Table S1. Search terms applied to each database

Database	Search Strategy
PubMed	("breast neoplasms"[MeSH Terms] OR ("breast"[All Fields] AND "neoplasms"[All Fields]) OR "breast neoplasms"[All Fields] OR ("breast"[All Fields] AND "cancer"[All Fields]) OR "breast cancer"[All Fields] OR "breast tumour"[All Fields] OR "breast malignancy"[All Fields]) AND (("polymorphism, single nucleotide"[MeSH Terms] OR ("polymorphism"[All Fields] AND "single"[All Fields] AND "nucleotide"[All Fields]) OR "single nucleotide polymorphism"[All Fields] OR ("single"[All Fields] AND "nucleotide"[All Fields] AND "polymorphism"[All Fields])) OR "snp"[All Fields] OR "snps"[All Fields] OR (common[All Fields] AND "genetic"[All Fields] AND variant[All Fields])) AND ((("risk"[MeSH Terms] OR "risk"[All Fields]) AND prediction[All Fields] AND model[All Fields]) OR (("risk"[MeSH Terms] OR "risk"[All Fields]) AND prediction[All Fields]) OR polygenic risk score)
EMBASE	'risk prediction model' OR 'risk prediction' AND ('single nucleotide polymorphism' OR 'common genetic variant' OR 'genetic variants' OR 'snp' OR 'single-nucleotide polymorphism' OR 'polygenic risk score' OR 'polygenetic risk score') AND ('breast cancer' OR 'breast tumour' OR 'breast malignan*' OR 'breast neoplasm*')
SCOPUS	(TITLE-ABS-KEY (**breast cancer**) OR TITLE-ABS-KEY (breast tumour) OR TITLE-ABS-KEY (breast malignan*)) AND TITLE-ABS-KEY (**risk prediction model**) AND (TITLE-ABS-KEY (**single nucleotide polymorphism***) OR TITLE-ABS-KEY (**polygenic risk score**) OR TITLE-ABS-KEY (**common genetic variant***) OR TITLE-ABS-KEY (polygenetic risk score))

Table S2a. 25-item "Strengthening the Reporting of Genetic RIsk Prediction Studies (GRIPS)" checklist*

TITLE & ABSTRACT		
	1	(a) Identify the article as a study of risk prediction using genetic factors. (b) Use recommended keywords in the abstract: genetic or genomic, risk, prediction.
INTRODUCTION		
Background and rationale	2	Explain the scientific background and rationale for the prediction study.
Objectives	3	Specify the study objectives and state the specific model(s) that is/are investigated. State if the study concerns the development of the model(s), a validation effort, or both.
METHODS		
Study design and setting	4*	Specify the key elements of the study design and describe the setting, locations, and relevant dates, including periods of recruitment, follow-up, and data collection.
Participants	5*	Describe eligibility criteria for participants, and sources and methods of selection of participants.
Variables: Definition	6*	Clearly define all participant characteristics, risk factors and outcomes. Clearly define genetic variants using a widely-used nomenclature system.
Variables: Assessment	7*	(a) Describe sources of data and details of methods of assessment (measurement) for each variable. (b) Give a detailed description of genotyping and other laboratory methods.
Variables: Coding	8	(a) Describe how genetic variants were handled in the analyses. (b) Explain how other quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why.
Analysis: Risk model construction	9	Specify the procedure and data used for the derivation of the risk model. Specify which candidate variables were initially examined or considered for inclusion in models. Include details of any variable selection procedures and other model-building issues. Specify the horizon of risk prediction (e.g., 5-year risk).
Analysis: Validation	10	Specify the procedure and data used for the validation of the risk model.
Analysis: Missing data	11	Specify how missing data were handled.
Analysis: Statistical methods	12	Specify all measures used for the evaluation of the risk model including, but not limited to, measures of model fit and predictive ability.
Analysis: Other	13	Describe all subgroups, interactions, and exploratory analyses that were examined.
RESULTS		
Participants	14*	Report the numbers of individuals at each stage of the study. Give reasons for nonparticipation at each stage. Report the number of participants not genotyped, and reasons why they were not genotyped.
Descriptives: Population	15*	Report demographic and clinical characteristics of the study population, including risk factors used in the risk modeling.
Descriptives: Model estimates	16	Report unadjusted associations between the variables in the risk model(s) and the outcome. Report adjusted estimates and their precision from the full risk model(s) for each variable.
Risk distributions	17*	Report distributions of predicted risks and/or risk scores.
Assessment	18	Report measures of model fit and predictive ability, and any other performance measures, if pertinent.
Validation	19	Report any validation of the risk model(s).
Other analyses	20	Present results of any subgroup, interaction, or exploratory analyses, whenever pertinent.
DISCUSSION		
Limitations	21	Discuss limitations and assumptions of the study, particularly those concerning study design, selection of participants, and measurements and analyses, and discuss their impact on the results of the study.
Interpretation	22	Give an overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.
Generalizability	23	Discuss the generalizability and, if pertinent, the health care relevance of the study results.
OTHER		
Supplementary information	24	State whether databases for the analyzed data, risk models, and/or protocols are or will become publicly available and if so, how they can be accessed.
Funding	25	Give the source of funding and the role of the funders for the present study. State whether there are any conflicts of interest.

* Marked items should be reported for every population in the study.
doi:10.1371/journal.pmed.1000420.t001

*Obtained from Janssens AC, Ioannidis JP, Bedrosian S, et al. Strengthening the reporting of genetic risk prediction studies (GRIPS): explanation and elaboration. Eur J Clin Invest. 2011;41(9):1010-1035.

Table S2b. Evaluation of studies against the 25-item "Strengthening the Reporting of Genetic RIsk Prediction Studies Prediction Studies" (GRIPS) checklist

Author, year/Item	1	2	3	4	5	6	7a	7 b	8a	8 b	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Mealiffe 2010																											
Wacholder 2010																		*									
Zheng 2010																		*									
Kaklamani 2011																		*									
Dai 2012																		*									
Darabi 2012																		*									
Higginbottham 2012																		*									
Hüsing 2012																		*									
Sueta 2012																		*									
Dite 2013																		*									
Xu 2013																											
Jupe 2014																											
Lee 2014																		*		**							
Lee 2015																											
Allman 2015[a]																		*									
Vachon 2015																		*									
Wu 2015																											
Burnside 2016																											
Dite 2016																											
Shieh 2016																		*									
Wen 2016[a]																		*									

Maas 2016																		*										
Hsieh 2017																		*										
Guo 2017																		*										
Shieh 2017																		*										
van Veen 2018																												
No. of items not reported	7	0	5	5	4	3	8	5	0	7	2	15	14	0	11	17	4	20	11	0	18	15	1	0	9	26	0	

White box indicates that the item was reported.

Gray box indicates that the item was not reported.

* indicates that Single Nucleotide Polymorphism were reported.

** Study provided NRI but not AUC.

Table S3a. Newcastle-Ottawa Quality Assessment Scale (NOS) coding manual for case-control studies*

SELECTION

1) Is the Case Definition Adequate?

a) Requires some independent validation (e.g. >1 person/record/time/process to extract information, or reference to primary record source such as x-rays or medical/hospital records) ☆

b) Record linkage (e.g. ICD codes in database) or self-report with no reference to primary record

c) No description

2) Representativeness of the Cases

a) All eligible cases with outcome of interest over a defined period of time, all cases in a defined catchment area, all cases in a defined hospital or clinic, group of hospitals, health maintenance organisation, or an appropriate sample of those cases (e.g. random sample) ☆

b) Not satisfying requirements in part (a), or not stated.

3) Selection of Controls

This item assesses whether the control series used in the study is derived from the same population as the cases and essentially would have been cases had the outcome been present.

a) Community controls (i.e. same community as cases and would be cases if had outcome) ☆

b) Hospital controls, within same community as cases (i.e. not another city) but derived from a hospitalised population

c) No description

4) Definition of Controls

a) If cases are first occurrence of outcome, then it must explicitly state that controls have no history of this outcome. If cases have new (not necessarily first) occurrence of outcome, then controls with previous occurrences of outcome of interest should not be excluded. ☆

b) No mention of history of outcome

COMPARABILITY

1) Comparability of Cases and Controls on the Basis of the Design or Analysis

- A maximum of 2 stars can be allotted in this category
- Either cases and controls must be matched in the design and/or confounders must be adjusted for in the analysis. Statements of no differences between groups or that differences were not statistically significant are not sufficient for establishing

comparability. Note: If the odds ratio for the exposure of interest is adjusted for the confounders listed, then the groups will be considered to be comparable on each variable used in the adjustment.
- There may be multiple ratings for this item for different categories of exposure (e.g. ever vs. never, current vs. previous or never)
Age = ☆, Other controlled factors = ☆

EXPOSURE

1) Ascertainment of exposure

a) Secure record (eg surgical records) ☆

b) Structured interview where blind to case/control status ☆

c) Interview not blinded to case/control status

d) Written self-report or medical record only

e) No description

2) Same method of ascertainment for cases and controls

a) Yes ☆

b) No

3) Non-Response rate

a) Same rate for both groups ☆

b) Non-respondents described

c) Rate different and no designation

*Obtained from GA Wells BS, D O'Connell, J Peterson, V Welch, M Losos, P Tugwell. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. <http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp>. Accessed 2018 1 April.

Table S3b. Newcastle-Ottawa Quality Assessment Scale (NOS) of Studies

Author, year	Selection				Comparability	Exposure			Total no. of stars
	Q1	Q2	Q3	Q4	Q1	Q1	Q2	Q3	
Mealiffe, 2010	*	*	*	*	**	*	*		8
Wacholder, 2010	*	*	*	*	*	*	*		7
Zheng, 2010	*	*	*	*	**	*	*		8
Kaklamani, 2011	*	*		*	**	*	*		7
Dai, 2012	*	*	*	*	**	*	*		8
Darabi, 2012		*	*		**		*		5
Higginbottham, 2012	*	*	*	*	*	*	*		7
Hüsing, 2012	*	*	*		**	*	*		7
Sueta, 2012		*			**	*	*		5
Dite, 2013	*	*	*	*	*	*	*		7
Xu, 2013	*	*	*	*		*	*		6
Jupe, 2014		*	*	*	*	*	*	*	7
Lee, 2014		*	*	*	**	*	*	*	8
Lee, 2015	*	*	*			*	*		5
Allman, 2015		*			*				3
Vachon, 2015	*	*		*	**		*		6
Wu, 2015		*		*	*		*		4
Burnside, 2016	*	*	*	*	*	*	*		7
Dite, 2016		*	*	*	*	*	*		6
Shieh, 2016		*		*	**	*	*		6
Wen, 2016					**				2
Maas, 2016	*			*	**	*			5
Hsieh, 2017	*	*	*	*	**	*	*		8
Guo, 2017	*	*	*		*		*		5
Shieh, 2017	*		*	*	**	*	*		7
van Veen, 2018	*		*	*		*	*		5

Table S4. Detailed characteristics of studies that evaluated the predictive ability of SNP-enhanced breast cancer risk prediction models

Author	Participants with BC	Participants with no BC	Detailed method of incorporating SNPs in model	Loci of SNPs used in model
Mealiffe, 2010	1,664	1,636	Mealiffe method - SNP relative score (SNP risk) was the product of adjusted risk values for each SNPs. Risk allele frequencies and published estimates of odds ratio per allele were used in the calculations based on a log-additive risk model	rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084, rs3817198
Wacholder, 2010	5,590	5,998	Allele count - In the genetic model, the risky alleles present (0, 1 or 2) were quantified for each individual SNP.	rs1045485, rs13281615, rs13387042, rs2981582, rs3803662, rs3817198, rs889312, rs7716600, rs11249433, rs999737
Zheng, 2010	3,039	3,082	Risk score - Genetic risk score was created by multiplying regression coefficient derived from a logistic model for each SNP replicated in the study individually	rs2046210, rs1219648, rs3817198, rs8051542, rs3803662, rs889312, rs10941679, rs13281615
Kaklamani, 2011	354	364	Other method - Multiple-SNP analyses used Bayesian hierarchical logistic models using prior distributions for genetic effects.	rs7206790, rs8047395, rs9939609, rs1477196
Dai, 2012	914	967	Allele count - 1. Each risk allele/factor treated equally and combined them based on counts of risk alleles/factors	rs13387042, rs2307032, rs2180341, rs2046210, rs2981582
			Risk score - 2. Risk score calculated with a linear combination of SNP alleles or risk factors weighted by individual odds ratio and grouped into quartiles	
			Modified Gail model - 3. Absolute risk for each woman estimated by modified Gail model and used a multiplicative model to derive genotype relative risk from the allelic odds ratio	
Darabi, 2012	1,566	1,527	Multiplicative penetrance model	rs11249433, rs1045485, rs13387042, rs4973768, rs10941679, rs889312, rs2046210, rs13281615, rs1011970, rs2981582, rs2380205, rs10995190, rs704010, rs3817198, rs614367, rs999737, rs3803662, rs6504950
	1,017	856		
	1,552	1,185		

Higginbotham, 2012	599	1,161	Other method - The combined effects on BC risk of SNP variant and proliferative disease in the patient's entry biopsy was included in the model as the genetic component.	well-establised BC risk loci: rs1219648, rs11200014, rs2420946, rs2981579, rs1626678, rs1681723, rs740363; Novel loci: rs1626678, rs8046508

Author	Participants with BC	Participants with no BC	Detailed method of incorporating SNPs in model	Loci of SNPs used in model
Hüsing, 2012	6,009	7,827	Allele count - In the genetic model, the risky alleles present (0, 1 or 2) were quantified for each individual SNP.	rs2981582, rs3803662, rs889312, rs3817198, rs13281615, rs13387042, rs1045485
				rs1045485, rs13281615, rs13387042, rs2981582, rs3803662, rs3817198, rs889312, rs11249433, rs999737
			Log-additive model - 18 SNPs: model obtained by construction- included SNPs into a multiple log-additive model with individually weighted per-allele effects for each SNP	rs11249433, rs1045485, rs13387042, rs4973768, rs10069690, rs10941679, rs889312, rs1562430, rs865686, rs2380205, rs10995190, rs1250003, rs2981582, rs614367, rs999737, rs3803662, rs6504950
			Allele count - In the genetic model, the risky alleles present (0, 1 or 2) were quantified for each individual SNP.	rs11249433, rs10931936, rs1045485, rs13387042, rs4973768, rs10069690, rs4415084, rs10941679, rs889312, rs2180341, rs9383935, rs3757318, rs9383938, rs2046210, rs13281615, rs1562430, rs1011970, rs865686, rs2380205, rs10995190, rs16917302, rs1250003, rs3750817, rs2981582, rs3817198, rs909116, rs614367, rs999737, rs3803662, rs2075555, rs6504950, rs311499
Sueta, 2012	697	1,394	Risk score - A PRS was created to measure the cumulative effect of multiple genetic risk variants	rs2981579, rs3803662, rs2046210, rs13281615, rs4973768, rs38137198, rs10931936
Dite, 2013	962	463	Mealiffe method - Using approach in Mealiffe 2010, SNP relative score (SNP risk) was the product of adjusted risk values for each SNPs. Risk allele frequencies and published estimates of odds ratio per allele were used in the calculations based on a log-additive risk model	rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084, rs3817198
Xu, 2013	298	612	Allele count - In the genetic model, the risky alleles present (0, 1 or 2) were quantified for each individual SNP.	rs13387042, rs889312, rs2180341, rs13281615, rs10736303, rs2981582, rs3817198, rs999737, rs3803662

Jupe, 2014	1,671	3,351	Other method - SNPs are weighted individually and/or by interaction with a specific age stratum in the absence/presence of an affected first-degree family member, in a polyfactorial risk model.	rs34915260 (Promoter PIII), rs34915260 (Exon 1), rs2252757, rs4680, rs1799998, rs10046, rs4646903, rs1800440, rs10012, rs1051740, rs17655, rs2077647, rs2000993, rs3842752, rs3745535, rs3136229, rs4796033, rs1799725, rs763110, rs7975232, rs2228000, rs3218536
Lee, 2014	411	1,212	Sum of product method - PRS was calculated by obtaining the product of number of risk alleles that the individual carries (0, 1, 2) and estimated log odds ratio for the association between SNP and BC, and then summing information from all SNPs.	rs616488, rs11552449, rs11249433, rs4849887, rs2016394, rs13387042, rs16857609, rs6762644, rs4973768, rs12493607, rs9790517, rs6828523, rs4415084, rs10941679, rs889312, rs10472076, rs1432679, rs11242675, rs2180341, rs204247, rs17529111, rs3757318, rs2046210, rs720475, rs9693444, rs13281615, rs1562430, rs1011970, rs10759243, rs865686, rs10822013, rs10995190, rs704010, rs1219648, rs2981582, rs3817198, rs3903072, rs11820646, rs12422552, rs10771399, rs17356907, rs2236007, rs941764, rs4784227, rs3112612, rs13329835, rs17817449, rs527616, rs1436904, rs3760982, rs2823093
Lee, 2015	680	23,481	Sum of product method - PRS was calculated by obtaining the product of number of risk alleles that the individual carries (0, 1, 2) and estimated log odds ratio for the association between SNP and BC, and then summing information from all SNPs.	rs12022378, rs11249433, rs616488, rs4245739, rs6678914, rs13387042, rs16857609, rs1045485, rs4849887, rs2016394, rs1550623, rs12710696, rs4973768, rs6762644, rs12493607, rs9790517, rs6828523, rs7726159, rs10069690, rs2736108, rs10941679, rs889312, rs10472076, rs1353747, rs1432679, rs12662670, rs2046210, rs11242675, rs204247, rs17529111, rs720475, rs13281615, rs9693444, rs6472903, rs2943559, rs11780156, rs10759243, rs865686, rs1011970, rs2981579, rs11199914, rs7072776, rs11814448, rs2380205, rs10995190, rs704010, rs7904519, rs554219, rs3817198, rs3903072, rs11820646, rs10771399, rs1292011, rs12422552, rs17356907, rs11571833, rs2588809, rs999737, rs2236007, rs941764, rs3803662, rs17817449, rs13329835, rs11075995, rs6504950, rs527616, rs1436904, rs8170, rs2363956, rs4808801, rs3760982, rs2823093, rs17879961, rs132390, rs6001930
Allman, 2015	416	7,005	Mealiffe method - Using approach in Mealiffe 2010, SNP relative score (SNP risk) was the product of adjusted risk values for each SNPs. Risk allele frequencies and published estimates of odds ratio per allele were used in the calculations based on a log-additive risk model	African American: rs11249433, rs4245739, rs616488, rs6678914, rs1045485, rs12710696, rs13387042, rs1550623, rs16857609, rs2016394, rs4849887, rs12493607, rs4973768, rs6762644, rs6828523, rs9790517, rs10069690, rs10472076, rs10941679, rs1353747, rs1432679, rs4415084, rs889312, rs11242675, rs17529111, rs204247, rs2046210, rs3757318, rs720475, rs11780156, rs13281615, rs2943559, rs6472903, rs9693444, rs1011970, rs10759243, rs865686, rs10995190, rs11199914, rs11814448, rs2380205, rs2981579, rs2981582, rs704010, rs7072776, rs7904519, rs11820646, rs3817198, rs3903072, rs554219, rs614367, rs75915166, rs10771399, rs12422552, rs1292011, rs17356907, rs11571833, rs2236007, rs2588809, rs941764, rs999737, rs11075995, rs13329835, rs17817449, rs3803662, rs6504950, rs1436904, rs527616, rs3760982, rs4808801, rs8170, rs2284378, rs2823093, rs132390, rs6001930

	147	3,201		Hispanics: rs11249433, rs11552449, rs4245739, rs616488, rs6678914, rs12710696, rs13387042, rs1550623, rs16857609, rs2016394, rs4849887, rs12493607, rs4973768, rs6762644, rs6828523, rs7696175, rs9790517, rs10069690, rs10472076, rs10941679, rs1353747, rs1432679, rs2067980, rs889312, rs11242675, rs140068132, rs204247, rs2046210, rs2180341, rs17157903, rs720475, rs11780156, rs13281615, rs2943559, rs6472903, rs9693444, rs1011970, rs10759243, rs865686, rs10995190, rs11199914, rs11814448, rs2380205, rs2981579, rs2981582, rs704010, rs7072776, rs11820646, rs3817198, rs3903072, rs10771399, rs12422552, rs1292011, rs17356907, rs2236007, rs2588809, rs941764, rs999737, rs11075995, rs13329835, rs17817449, rs3803662, rs6504950, rs1436904, rs527616, rs2363956, rs3760982, rs4808801, rs8170, rs2823093, rs6001930

Author	Participants with BC	Participants with no BC	Detailed method of incorporating SNPs in model	Loci of SNPs used in model
Vachon, 2015	334	334	Sum of product method - PRS was calculated by obtaining the product of number of risk alleles that the individual carries (0, 1, 2) and estimated log odds ratio for the association between SNP and BC, and then summing information from all SNPs.	rs616488, rs11552449 (proxy: rs3671936), rs11249433, rs6678914, rs4245739, rs12710696, rs4849887, rs2016394, rs1550623, rs1045485, rs13387042, rs16857609, rs6762644, rs4973768, rs12493607, rs9790517, rs6828523, rs10069690, rs2736108, rs10941679, rs889312, rs10472076, rs1353747, rs1432679, rs11242675, rs204247, rs17529111 (proxy: rs17530068), rs3757318, rs2046210, rs720475, rs9693444, rs6472903, rs2943559, rs13281615, rs11780156, rs1011970, rs10759243, rs865686, rs2380205, rs7072776, rs11814448, rs10995190, rs704010, rs7904519, rs11199914, rs2981579, rs2981582, rs3817198, rs3903072, rs614367, rs554219, rs11820646, rs75915166, rs12422552, rs10771399, rs17356907, rs1292011, rs11571833, rs2236007, rs2588809, rs999737, rs941764, rs3803662, rs17817449, rs11075995, rs13329835, rs6504950, rs527616, rs1436904, rs8170, rs2363956, rs4808801, rs3760982, rs2823093, rs132390, rs6001930 (proxy: rs6001913)
Wu, 2015	373	395	Allele count - In the genetic model, the risky alleles present (0, 1 or 2) were quantified for each individual SNP.	rs1045485, rs13281615, rs13387042, rs2981582, rs3803662, rs3817198, rs889312, rs7716600, rs11249433, rs999737
Wen, 2016	11,905	11,662	Sum of product method - PRS was calculated by obtaining the product of number of risk alleles that the individual carries (0, 1, 2) and estimated log odds ratio for the association between SNP and BC, and then summing information from all SNPs.	rs616488, rs11249433, rs4951011, rs12710696, rs4849887, rs10931936, rs13387042, rs16857609, rs4973768, rs12493607, rs6828523, rs10069690, rs10941679, rs889312, rs10474352, rs1432679, rs9485372, rs2046210, rs9693444, rs6472903, rs1562430, rs1011970, rs10759243, rs10822013, rs704010, rs11199914, rs2981579, rs909116, rs614367, rs7107217, rs12422552, rs10771399, rs17356907, rs1292011, rs2236007, rs941764, rs2290203, rs3803662, rs4784227, rs11075995, rs527616, rs2363956, rs4808801, rs12628403
Burnside, 2016	373	395	Allele count - In the genetic model, the risky alleles present (0, 1 or 2) were quantified for each individual SNP.	rs1045485, rs13281615, rs13387042, rs2981582, rs3803662, rs3817198, rs889312, rs10941679, rs999737, rs11249433

Dite, 2016	405	750	Mealiffe method - SNP relative score (SNP risk) was the product of adjusted risk values for each SNPs. Risk allele frequencies and published estimates of odds ratio per allele were used in the calculations based on a log-additive risk model	rs616488, rs11552449, rs11249433, rs6678914, rs4245739, rs12710696, rs4849887, rs2016394, rs1550623, rs1045485, rs13387042, rs16857609, rs6762644, rs4973768, rs12493607, rs9790517, rs6828523, rs10069690, rs7726159, rs2736108, rs10941679, rs889312, rs10472076, rs1353747, rs1432679, rs11242675, rs204247, rs17529111, rs12662670, rs2046210, rs720475, rs9693444, rs6472903, rs2943559, rs13281615, rs11780156, rs1011970, rs10759243, rs865686, rs2380205, rs7072776, rs11814448, rs10995190, rs704010, rs7904519, rs11199914, rs2981579, rs3817198, rs3903072, rs78540526, rs554219, rs75915166, rs11820646, rs12422552, rs10771399, rs17356907, rs1292011, rs11571833, rs2236007, rs2588809, rs999737, rs941764, rs3803662, rs17817449, rs11075995, rs13329835, rs6504950, rs527616, rs1436904, rs8170, rs2363956, rs4808801, rs3760982, rs2823093, rs17879961, rs132390, rs6001930
Shieh, 2016	471	460	Bayesian method - PRS was calculated using allele frequencies and odds ratios from Caucasian population, using a Bayesian approach, as the composite likelihood ratio representing the individual effects of each SNP	rs10069690, rs1011970, rs10472076, rs10474352, rs10759243, rs10771399, rs10822013, rs10941679, rs10995190, rs11075995, rs11199914, rs11242675, rs11249433, rs11552449, rs11571833, rs11780156, rs11814448, rs11820646, rs12422552, rs12493607, rs12710696, rs1292011, rs132390, rs13281615, rs13329835, rs13387042, rs1353747, rs140068132, rs1432679, rs1436904, rs1550623, rs1562430, rs16857609, rs17356907, rs17530068, rs17817449, rs17879961, rs204247, rs2046210, rs2236007, rs2284378, rs2290203, rs2380205, rs2392780, rs2588809, rs2736108, rs2823093, rs2943559, rs2981579, rs3757318, rs3760982, rs3803662, rs3817198, rs3822625, rs3903072, rs4245739, rs4415084, rs4808801, rs4849887, rs4951011, rs4973768, rs527616, rs554219, rs6001930, rs614367, rs616488, rs6472903, rs6504950, rs6678914, rs6762644, rs6828523, rs704010, rs7072776, rs720475, rs75915166, rs7726354, rs7904519, rs8170, rs865686, rs889312, rs9383938, rs941764, rs9485372, rs9693444, rs9790517, rs999737

Author	Participants with BC	Participants with no BC	Detailed method of incorporating SNPs in model	Loci of SNPs used in model
Maas, 2016	17,171	19,862	Sum of product method - PRS was calculated by obtaining the product of number of risk alleles that the individual carries (0, 1, 2) and estimated log odds ratio for the association between SNP and BC, and then summing information from all SNPs.	24 SNPs genotyped in BPC3: rs11249433, rs1045485, rs13387042, rs4973768, rs10069690, rs10941679, rs889312, rs17530068, rs2046210, rs1562430, rs1011970, rs865686, rs2380205, rs10995190, rs1250003, rs2981582, rs909116, rs614367, rs10483813, rs3803662, rs6504950, rs8170, rs2284378, rs999737; 68 SNPs: rs75915166, rs554219, rs2736108, rs2588809, rs10759243, rs11199914, rs7072776, rs11814448, rs16857609, rs11552449, rs12662670, rs10771399, rs1292011, rs2363956, rs2823093, rs17879961, rs616488, rs4849887, rs2016394, rs1550623, rs6762644, rs12493607, rs9790517, rs6828523, rs10472076, rs1353747, rs1432679, rs11242675, rs204247, rs720475, rs9693444, rs6472903, rs2943559, rs11780156, rs7904519, rs11820646, rs12422552, rs17356907, rs11571833, rs2236007, rs941764, rs17817449, rs13329835, rs527616, rs1436904, rs4808801, rs3760982, rs132390, rs6001930, rs4245739, rs6678914, rs12710696, rs11075995, rs12405132, rs12048493, rs72755295, rs6796502, rs13162653, rs2012709, rs7707921, rs9257408, rs4593472, rs13365225, rs13267382, rs11627032, chr17:29230520:D, rs745570, rs6507583
Hsieh, 2017	446	514	Sum of product method - PRS was calculated by obtaining the product of number of risk alleles that	rs2981579, rs3750817, rs1219648, rs2981582, rs981782, rs889312, rs1686165, rs3803662, rs4784227, rs2046210, rs10822013, rs3784099, rs13393577

			the individual carries (0, 1, 2) and estimated log odds ratio for the association between SNP and BC, and then summing information from all SNPs.	
Guo, 2017	103	303	Other method - Co-dominant model, dominant model and recessive model of risk alleles were established and if showed significant association with risk, were included in the multivariate binary logistic regression analyses with the other risk factors.	rs3803662, rs2046210, rs4784227, rs4973768
	150	602		rs2046210, rs4784227
Shieh, 2017	110	214	Bayesian method - PRS was calculated using allele frequencies and odds ratios from Caucasian population, using a Bayesian approach, as the composite likelihood ratio representing the individual effects of each SNP	rs10069690, rs1011970, rs10472076, rs10474352, rs10759243, rs10771399, rs10822013, rs10941679, rs10995190, rs11075995, rs11199914, rs11242675, rs11249433, rs11552449, rs11571833, rs11780156, rs11814448, rs11820646, rs12422552, rs12493607, rs12710696, rs1292011, rs132390, rs13281615, rs13329835, rs13387042, rs1353747, rs140068132, rs1432679, rs1436904, rs1550623, rs1562430, rs16857609, rs17356907, rs17530068, rs17817449, rs17879961, rs204247, rs2046210, rs2236007, rs2284378, rs2290203, rs2380205, rs2392780, rs2588809, rs2736108, rs2823093, rs2943559, rs2981579, rs3757318, rs3760982, rs3803662, rs3817198, rs3822625, rs3903072, rs4245739, rs4415084, rs4808801, rs4849887, rs4951011, rs4973768, rs527616, rs554219, rs6001930, rs614367, rs616488, rs6472903, rs6504950, rs6678914, rs6762644, rs6828523, rs704010, rs7072776, rs720475, rs75915166, rs7726354, rs7904519, rs8170, rs865686, rs889312, rs9383938, rs941764, rs9485372, rs9693444, rs9790517, rs999737
van Veen, 2018	466	8,897	Risk score - PRS was calculated by multiplying the per-allele odds ratio for each SNP and normalizing the risk by the average risk expected in the population using published minor allele frequencies.	rs614367, rs704010, rs713588, rs889312, rs909116, rs1011970, rs1156287, rs1562430, rs2981579, rs3757318, rs3803662, rs4973768, rs8009944, rs9790879, rs10995190, rs11249433, rs13387042, rs10931936

BC = Breast cancer, SNP = Single nucleotide polymorphism, PRS = Polygenic risk score

Table S5. Improvement in AUC in SNP-enhanced breast cancer risk prediction model of, grouped by baseline model

Baseline model	No. of risk factors	No. of SNPs	Method of incorporating SNPs in model	AUC of baseline model (95% CI)	AUC of SNP-enhanced model (95% CI)	Increase in AUC (95% CI)	Author	Year
(A) Studies using BRCAT as baseline model								
BCRAT (5-6 risk factors of BCRAT)	5	7	Mealiffe method	0.58 (0.55, 0.61)	overall: 0.61 (0.58, 0.64)	0.03	Dite	2013
	5*	77	Mealiffe method	0.64 (0.60, 0.68)	0.67 (0.63, 0.70)	0.03	Dite	2016
	6	7	Mealiffe method	0.557 (0.547, 0.575)	0.594 (0.575, 0.612)	0.037 (0.025, 0.051)	Mealiffe	2010
	6	9	Other method	0.573	0.601	0.028	Higginbottham	2012
	6	75	Mealiffe method	0.56 (0.53, 0.59)	0.59 (0.56, 0.61)	0.03	Allman	2015
	6	71	Mealiffe method	0.55 (0.51, 0.60)	0.61 (0.56, 0.66)	0.06	Allman	2015
	6 + 2	75	Sum of product method	0.660 (0.640, 0.680)	0.680 (0.660, 0.690)	0.020	Lee	2015
Partial BCRAT (2-4 risk factors of BCRAT)	2	5	Risk score	0.638	Risk score: 0.649 (0.631, 0.667)	0.011	Dai	2012
	4	10	Allele count	0.580	0.618	0.038	Wacholder	2010
	4	9	Allele count	0.560	0.612	0.052	Xu	2013
	4	22	Other method	0.580	35-39 years: 0.69	0.11	Jupe	2014
Partial BCRAT + additional risk factors	2+2	18	Multiplicative penetrance model	0.548 (0.527, 0.568)	0.615 (0.595, 0.634)	0.067	Darabi	2012
	2+3	13	Sum of product method	0.634	0.665	0.031	Hsieh	2017
	4+2	9	Other method	0.588	0.615	0.027	Higginbottham	2012
	2+5	18	Multiplicative penetrance model	0.604 (0.579, 0.630)	0.619 (0.594, 0.644)	0.015	Darabi	2012
	2+5	8	Risk score	0.618	c-statistic: 0.630	c-statistic: 0.012	Zheng	2010
	3+5	7	Risk score	0.665	c-statistic: 0.693	c-statistic: 0.028	Sueta	2012
	3+7	10	Allele count	0.597	0.638	0.041	Wu	2015

	2+3	44	Sum of product method	0.563	0.602	0.039 (0.026, 0.051)	Wen	2016
	2+9	92	Sum of product method	0.588	0.648	0.06	Maas	2016
	3+11	4	Other method	0.708	0.785	0.077	Guo	2017
		2		0.693	0.764	0.071		

(B) Studies using BCSC, IBIS, BRCAPRO, BOADICEA, BCSC, IBIS and other models as baseline model

Baseline model	**No. of risk factors**	**No. of SNPs**	**Method of incorporating SNPs in model**	**AUC of baseline model**	**AUC of SNP-enhanced model**	**Increase in AUC**	**Author**	**Year**
BCSC	2	83	Bayesian method	0.62 (0.52, 0.73)	0.65 (0.61, 0.68)	0.03	Shieh	2016
	5	76	Sum of product method	0.66 (0.61, 0.70)	0.69 (0.67, 0.71)	0.03 (0.007, 0.049)	Vachon	2015
	6	83	Bayesian method	0.58 (0.50, 0.65)	0.64 (0.57, 0.71)	0.06	Shieh	2017
	6+1			0.67 (0.60, 0.74)	0.72 (0.65, 0.79)	0.05		
IBIS	12*	77	Mealiffe method	0.57 (0.53, 0.60)	0.63 (0.59, 0.66)	0.06	Dite	2016
	14+1*	18	Risk score	0.64 (0.60, 0.68)	0.67 (0.62, 0.71)	0.03	van Veen	2018
	6	75	Mealiffe method	0.51 (0.48, 0.54)	0.55 (0.52, 0.58)	0.04	Allman	2015
		71		0.53 (0.48, 0.57)	0.59 (0.54, 0.64)	0.06		
Personal information + Demographic factors	1	10	Allele count	0.547 (0.506, 0.587)	0.601 (0.562, 0.641)	0.054	Burnside	2016
	3	4	Other method	0.53	Epi: 0.68	Epi: 0.15	Kaklamani	2011
BOADICEA	8	77	Mealiffe method	0.66 (0.63, 0.70)	0.70 (0.67, 0.73)	0.04	Dite	2016
BRCAPRO	8	77		0.65 (0.62, 0.68)	0.69 (0.66, 0.72)	0.04	Dite	2016
Risk factors	10	7	Allele count	0.564 (0.547, 0.581)	0.591 (0.574. 0.608)	0.027	Hüsing	2012
		9			0.595 (0.579, 0.612)	0.031		
		18			SNPs in log-additive model: 0.605 (0.589, 0.622)	0.041		

32		0.604 (0.588, 0.621)	0.040

*Number of risk factors assumed to be all the risk factors used in the stated baseline model, excluding risk factors that were stated to be missing/no information.

Table S6. Risk factors included in the various breast cancer risk prediction models

Risk Factor	BCRAT	BCSC v1.0	BCSC v2.0	BRCAPRO[a]	IBIS[a]	BOADICEA
Personal Information						
Age	Yes	Yes	Yes	Yes	Yes	Yes
BMI	No	No	No	No	Yes	No
Race	Yes	Yes	Yes	Yes	No	No
Height	No	No	No	No	Yes	No
Hormonal and Reproductive Factors						
Age at menarche	Yes	No	No	No	Yes	No
Age at first live birth	Yes	No	No	No	Yes	No
Age at menopause	No	No	No	No	Yes	No
Parity	No	No	No	No	Yes	No
Personal history of breast disease						
Breast biopsies	Yes	Yes	Yes	No	No	No
Atypical ductal hyperplasia	Yes	No	No	No	Yes	No
Lobular carcinoma in situ	Yes	No	No	No	Yes	No
Breast density	No	Yes	Yes	No	No	No
Benign breast disease	No	No	Yes	No	No	No
Family History of breast and/or ovarian cancer						
First-degree relatives with breast cancer	Yes	Yes	Yes	Yes	Yes	Yes
Second-degree relatives* with breast cancer	No	No	No	Yes	Yes	Yes
Age of onset of breast cancer in a relative	No	No	No	Yes	Yes	Yes
Bilateral breast cancer in a relative	No	No	No	Yes	Yes	Yes
Ovarian cancer in a relative	No	No	No	Yes	Yes	Yes

BRCAT = Breast Cancer Risk Assessment Tool (6); BCSC = Breast Cancer Surveillance Consortium, v1.0 (11), v2.0 (12); BRCAPRO (8); IBIS = International Breast Intervention Study (7); BOADICEA = Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (9,10); BMI = Body Mass Index

This table was adapted from Amir et al (81).

[a] BRCAPRO was designed as a pretest for BRCA mutation (77) while IBIS was designed for use in high-risk population (66, 69).

*Both maternal and paternal relatives are included.

Figure S7: Meta-analysis of the overall area under the curve (AUC) for SNP-enhanced breast cancer risk prediction models by increasing baseline AUC and baseline model group (a) BCRAT (5-6 risk factors of BCRAT) (b) Partial BCRAT (2-4 risk factors of BCRAT) (c) Partial BCRAT + additional risk factors (d) BCSC (e) IBIS (f) Other models

A forest plot for the overall AUC and increasing baseline AUC by baseline model group is shown. The 95% CIs are denoted by the black tails originating from the black square, which is the point estimate for overall AUC in each model. The combined overall AUC estimate within each baseline model group is represented by the black diamond, where the diamond width corresponds to the 95% CI bounds. Heterogeneity was assessed using I^2-statistic to determine the extent of variation between study population estimates. Random-effects model is applied for all model groups except BCSC (I^2-statistic < 50%).

Abbreviations used in Figure S7:

AUC = Area under receiver operating characteristics curve, CI = 95% Confidence Interval, SNP = Single Nucleotide Polymorphisms, BCRAT = Breast Cancer Risk Assessment Tool (6), BCSC = Breast Cancer Surveillance Consortium (11,12), BOADICEA = Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (9,10), BRCAPRO (8), IBIS = International Breast Intervention Study (7)

Figure S7

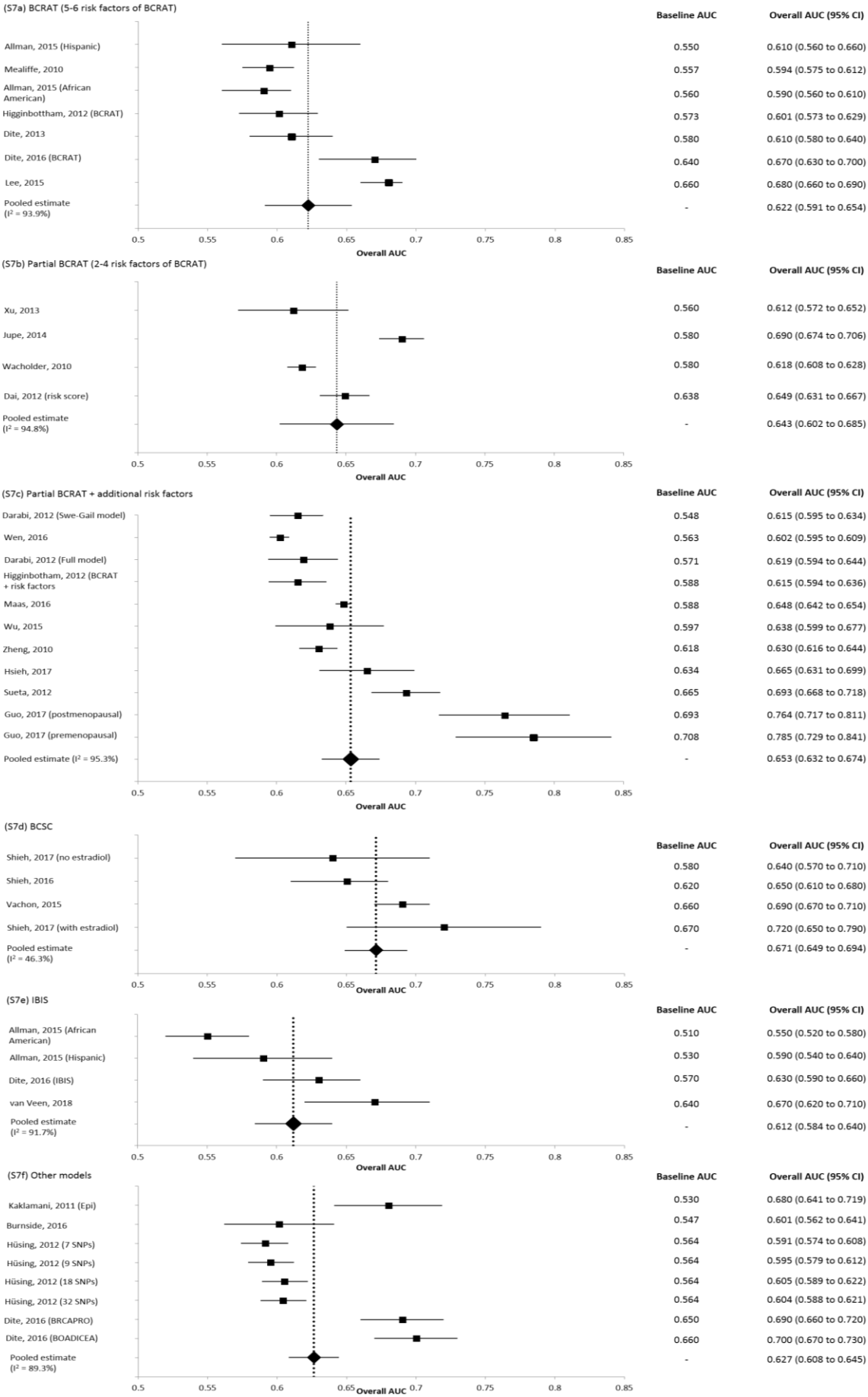
(S7a) BCRAT (5-6 risk factors of BCRAT)
Baseline AUC
Overall AUC (95% CI)
Allman, 2015 (Hispanic) 0.550 0.610 (0.560 to 0.660)
Mealiffe, 2010 0.557 0.594 (0.575 to 0.612)
Allman, 2015 (African American) 0.560 0.590 (0.560 to 0.610)
Higginbottham, 2012 (BCRAT) 0.573 0.601 (0.573 to 0.629)
Dite, 2013 0.580 0.610 (0.580 to 0.640)
Dite, 2016 (BCRAT) 0.640 0.670 (0.630 to 0.700)
Lee, 2015 0.660 0.680 (0.660 to 0.690)
Pooled estimate (I² = 93.9%) - 0.622 (0.591 to 0.654)
0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85
Overall AUC
(S7b) Partial BCRAT (2-4 risk factors of BCRAT)
Baseline AUC
Overall AUC (95% CI)
Xu, 2013 0.560 0.612 (0.572 to 0.652)
Jupe, 2014 0.580 0.690 (0.674 to 0.706)
Wacholder, 2010 0.580 0.618 (0.608 to 0.628)
Dai, 2012 (risk score) 0.638 0.649 (0.631 to 0.667)
Pooled estimate (I² = 94.8%) - 0.643 (0.602 to 0.685)
0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85
Overall AUC
(S7c) Partial BCRAT + additional risk factors
Baseline AUC
Overall AUC (95% CI)
Darabi, 2012 (Swe-Gail model) 0.548 0.615 (0.595 to 0.634)
Wen, 2016 0.563 0.602 (0.595 to 0.609)
Darabi, 2012 (Full model) 0.571 0.619 (0.594 to 0.644)
Higginbotham, 2012 (BCRAT + risk factors 0.588 0.615 (0.594 to 0.636)
Maas, 2016 0.588 0.648 (0.642 to 0.654)
Wu, 2015 0.597 0.638 (0.599 to 0.677)
Zheng, 2010 0.618 0.630 (0.616 to 0.644)
Hsieh, 2017 0.634 0.665 (0.631 to 0.699)
Sueta, 2012 0.665 0.693 (0.668 to 0.718)
Guo, 2017 (postmenopausal) 0.693 0.764 (0.717 to 0.811)
Guo, 2017 (premenopausal) 0.708 0.785 (0.729 to 0.841)
Pooled estimate (I² = 95.3%) - 0.653 (0.632 to 0.674)
0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85
Overall AUC
(S7d) BCSC
Baseline AUC
Overall AUC (95% CI)
Shieh, 2017 (no estradiol) 0.580 0.640 (0.570 to 0.710)
Shieh, 2016 0.620 0.650 (0.610 to 0.680)
Vachon, 2015 0.660 0.690 (0.670 to 0.710)
Shieh, 2017 (with estradiol) 0.670 0.720 (0.650 to 0.790)
Pooled estimate (I² = 46.3%) - 0.671 (0.649 to 0.694)
0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85
Overall AUC
(S7e) IBIS
Baseline AUC
Overall AUC (95% CI)
Allman, 2015 (African American) 0.510 0.550 (0.520 to 0.580)
Allman, 2015 (Hispanic) 0.530 0.590 (0.540 to 0.640)
Dite, 2016 (IBIS) 0.570 0.630 (0.590 to 0.660)
van Veen, 2018 0.640 0.670 (0.620 to 0.710)
Pooled estimate (I² = 91.7%) - 0.612 (0.584 to 0.640)
0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85
Overall AUC
(S7f) Other models
Baseline AUC
Overall AUC (95% CI)
Kaklamani, 2011 (Epi) 0.530 0.680 (0.641 to 0.719)
Burnside, 2016 0.547 0.601 (0.562 to 0.641)
Hüsing, 2012 (7 SNPs) 0.564 0.591 (0.574 to 0.608)
Hüsing, 2012 (9 SNPs) 0.564 0.595 (0.579 to 0.612)
Hüsing, 2012 (18 SNPs) 0.564 0.605 (0.589 to 0.622)
Hüsing, 2012 (32 SNPs) 0.564 0.604 (0.588 to 0.621)
Dite, 2016 (BRCAPRO) 0.650 0.690 (0.660 to 0.720)
Dite, 2016 (BOADICEA) 0.660 0.700 (0.670 to 0.730)
Pooled estimate (I² = 89.3%) - 0.627 (0.608 to 0.645)
0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85
Overall AUC

Figure S8: Summary of pooled estimates obtained from meta-analysis, grouped by similar baseline model

A summary of the pooled overall AUC estimates within each baseline model group is shown. The overall AUC estimate within each group is represented by the black diamond, where the diamond width corresponds to the 95% CI bounds.