

Additional file 1. Comparison of Age-adjusted RR (95% CI) for breast cancer among women in urban and rural areas of China Kadoorie Biobank

	Urban			Rural		
	Cases	Cases/PYs (/10,000)	RR (95% CI)	Cases	Cases/PYs (/10,000)	RR (95% CI)
Highest education						
No formal school	126	54.98	1.00 (reference)	213	40.17	1.00 (reference)
Primary school	257	96.41	1.17 (1.10 to 1.24)	313	45.99	1.17 (1.07 to 1.29)
Middle school	454	113.97	1.36 (1.21 to 1.53)	199	53.57	1.37 (1.14 to 1.65)
High school	448	145.2	1.59 (1.34 to 1.89)	57	58.71	1.61 (1.22 to 2.12)
College/university	206	168.25	1.85 (1.47 to 2.33)	14	130.35	1.89 (1.31 to 2.73)
BMI at attained age < 50 years, kg/m ²						
<18.5	10	111.76	1.00 (reference)	11	89.43	1.00 (reference)
<24.0	217	198.34	0.92 (0.79 to 1.07)	140	80.93	0.99 (0.83 to 1.19)
<28.0	109	257.99	0.84 (0.62 to 1.14)	78	104.95	0.99 (0.69 to 1.42)
≥28.0	20	169.83	0.77 (0.48 to 1.22)	23	115.65	0.98 (0.57 to 1.69)
BMI at attained age ≥ 50 years, kg/m ²						
<18.5	18	57.96	1.00 (reference)	18	29.43	1.00 (reference)
<24.0	405	101.48	1.25 (1.16 to 1.36)	237	42.02	1.25 (1.12 to 1.40)
<28.0	470	127.47	1.57 (1.34 to 1.84)	213	60.28	1.57 (1.25 to 1.96)
≥28.0	242	158.83	1.97 (1.56 to 2.49)	76	65.23	1.96 (1.40 to 2.74)
Height, cm						
100-	212	84.17	1.00 (reference)	170	34.79	1.00 (reference)
150.2-	299	95.9	1.14 (1.08 to 1.20)	205	46.14	1.13 (1.05 to 1.21)
154.2-	379	109.28	1.29 (1.16 to 1.43)	217	52.67	1.27 (1.10 to 1.46)
158.2-	601	144.97	1.47 (1.25 to 1.72)	204	59.11	1.43 (1.16 to 1.76)
Family history of cancer						
no family history	1,161	108.93	1.00 (reference)	634	44.03	1.00 (reference)
One relative history	270	120.62	1.04 (0.91 to 1.19)	132	60.19	1.24 (1.02 to 1.50)
More than one	60	169.7	1.42 (1.09 to 1.85)	30	97.48	1.96 (1.35 to 2.85)
Live birth count						
Nulliparous	36	141.05	1.86 (1.27 to 2.71)	9	62.61	1.74 (0.89 to 3.41)
1	856	125.38	1.77 (1.42 to 2.21)	211	57.8	1.49 (1.13 to 1.95)
2	366	117.81	1.43 (1.18 to 1.74)	353	53.61	1.38 (1.12 to 1.68)
≥3	233	76.11	1.00 (reference)	223	34.19	1.00 (reference)
Age at menarche, years						
<12	134	160.4	1.50 (1.24 to 1.80)	53	66.67	1.56 (1.21 to 2.00)
13-14	507	126.34	1.31 (1.16 to 1.48)	241	53.85	1.34 (1.14 to 1.59)
15-16	537	112.23	1.14 (1.08 to 1.22)	300	46.27	1.16 (1.07 to 1.26)
≥17	313	86.53	1.00 (reference)	202	39.25	1.00 (reference)

Abbreviations: BMI, body mass index; PY, person-year; RR, relative risk; CI, confidence interval.

Cox model was stratified by 5 study sites and age at baseline in a 5-year interval, and included all the predictors in the table.

Additional file 2. Age- and residence-specific breast cancer incidence rates and mortality rates of non-breast cancer per 100,000 person-years by data sources

	CKB				Nation				Shanghai	
	Urban		Rural		Urban		Rural		Urban	
	h_1	h_2	h_1	h_2	h_1	h_2	h_1	h_2	h_1	h_2
30-34	98.4	0.0	0.0	63.9	16.7	35.5	13.2	52.2	10.2	44.3
35-39	48.5	31.6	22.4	116.8	34.2	43.2	27.1	62.1	31.1	65.3
40-44	82.0	71.5	38.1	101.7	64.8	80.2	54.1	96.8	69.7	83.1
45-49	113.9	100.0	52.1	170.5	85.7	101.1	70.8	134.7	105	128.1
50-54	122.1	160.6	57.2	250.8	105.2	225.6	71.5	268.5	115.2	203.1
55-59	123.2	222.9	53.7	427.5	107.7	275.9	73.6	344.3	109.9	317.5
60-64	138.6	382.0	52.3	690.5	108.6	606.4	73.2	683.5	106.6	530.9
65-69	118.7	687.1	40.4	1208.0	98.7	1105.4	52.3	1285.2	103.3	993.7
70-74	100.6	1274.6	38.0	2152.1	92.5	1843.2	46.0	2225.7	113.3	1838.1
75-79	100.5	2422.5	34.6	3879.7	83.4	3482.2	39.0	3821.9	106	3547.3
80-84	81.7	4346.7	21.7	6699.8	74.9	6984.1	30.0	6776.3	74.9	6984.1
85-89	103.7	7925.7	157.9	7890.3	47.8	16282.3	22.1	14950.6	47.8	16282.3

Abbreviation: CKB, China Kadoorie Biobank; h_1 , incidence rate of breast cancer; h_2 , mortality rates of non-breast cancer.

CKB's rates were estimated among participants included in model derivation (2004–2016, n=300,824). National rates were calculated using incidence rates and breast cancer-specific mortality rates in 2014 from National Central Cancer Registry of China and all-cause mortality rates in 2014 from Health Statistics Yearbook. Shanghai's rates in age groups 30-79 were obtained from Zheng W, Wen W, Gao YT, et al. J Natl Cancer Inst 2010;102:972-81, while the rates in age groups 80-89 were replaced with national rates as they are not available in Zheng W et al.

Additional file 3. Age- and site-adjusted RR (95% CI) from the derivation subcohort and the whole China Kadoorie Biobank

	Derivation subcohort			Whole cohort		
	Cases	Cases/PYs (/100,000)	RR (95% CI)	Cases	Cases/PYs (/100,000)	RR (95% CI)
Highest education						
No formal school	222	43.65	1.00 (Reference)	339	44.64	1.00 (Reference)
Primary school	388	61.47	1.19 (1.12 to 1.26)	570	60.18	1.17 (1.11 to 1.23)
Middle school	441	86.19	1.41 (1.25 to 1.58)	653	84.82	1.37 (1.24 to 1.51)
High school	339	125.95	1.67 (1.40 to 1.99)	505	124.5	1.60 (1.39 to 1.85)
College/university	146	163.88	1.98 (1.56 to 2.50)	220	165.2	1.87 (1.54 to 2.27)
BMI at age < 50 years, kg/m ²						
<18.5	17	118.09	1.00 (Reference)	21	98.83	1.00 (Reference)
18.5-23.9	235	124.68	0.97 (0.84 to 1.12)	357	126.42	0.95 (0.84 to 1.07)
24.0-27.9	116	148.92	0.95 (0.71 to 1.26)	187	160.42	0.90 (0.71 to 1.14)
≥28	35	164.52	0.92 (0.60 to 1.41)	43	135.81	0.85 (0.60 to 1.21)
BMI at age ≥ 50 years, kg/m ²						
<18.5	23	39.28	1.00 (Reference)	36	41.07	1.00 (Reference)
18.5-23.9	438	68.05	1.23 (1.14 to 1.33)	642	66.66	1.25 (1.17 to 1.33)
24.0-27.9	472	98.43	1.51 (1.29 to 1.77)	683	94.59	1.57 (1.38 to 1.78)
≥28	200	111.94	1.86 (1.47 to 2.35)	318	118.27	1.96 (1.62 to 2.38)
Height, cm						
<150.2	260	52.68	1.00 (Reference)	382	51.58	1.00 (Reference)
150.2-154.1	353	69.95	1.12 (1.06 to 1.18)	504	66.66	1.13 (1.09 to 1.18)
154.2-158.1	379	74.94	1.25 (1.13 to 1.38)	596	78.54	1.28 (1.18 to 1.39)
≥158.2	544	107.56	1.40 (1.20 to 1.63)	805	105.96	1.45 (1.28 to 1.65)
No. of first-degree relatives diagnosed with overall cancer						
0	1,222	73.2	1.00 (Reference)	1,795	71.63	1.00 (Reference)
1	258	87.03	1.05 (0.91 to 1.20)	402	90.71	1.10 (0.99 to 1.23)
≥2	56	128.2	1.46 (1.11 to 1.91)	90	136.09	1.57 (1.27 to 1.95)
No. of live birth						
Nulliparous	33	124.9	1.84 (1.26 to 2.68)	45	112.79	1.78 (1.29 to 2.45)
1	714	102.53	1.57 (1.28 to 1.93)	1,067	101.84	1.66 (1.40 to 1.96)
2	476	73.41	1.32 (1.12 to 1.57)	719	74.19	1.41 (1.22 to 1.62)
≥3	313	49.03	1.00 (Reference)	456	47.58	1.00 (Reference)
Age at menarche, years						
<12	127	117.02	1.39 (1.16 to 1.67)	187	114.7	1.52 (1.31 to 1.76)
13-14	482	85.18	1.25 (1.10 to 1.41)	748	88.12	1.32 (1.20 to 1.46)
15-16	562	74.85	1.12 (1.05 to 1.19)	837	74.27	1.15 (1.09 to 1.21)
≥17	365	62.46	1.00 (Reference)	515	58.77	1.00 (Reference)

Abbreviations: BMI, body mass index; PY, person-year; RR, relative risk; CI, confidence interval.

Cox model was stratified by study sites (10 areas) and age at baseline in 5-year interval, and included all the predictors in the table.

Additional file 4. Expected and observed number of breast cancer in the test subcohort of China Kadoorie Biobank (CKB) and Shanghai Women's Health Study (SWHS) using the corresponding local rates

	Test subcohort of CKB ^a			SWHS ^b		
	<i>E</i>	<i>O</i>	<i>E/O</i> (95% CI)	<i>E</i>	<i>O</i>	<i>E/O</i> (95% CI)
Overall	765	751	1.02 (0.95-1.10)	1,423	1,409	1.01 (0.96-1.06)
Age at enrollment, years						
<50	372	357	1.04 (0.94-1.16)	793	688	1.15 (1.07-1.24)
50-59	278	256	1.08 (0.96-1.23)	391	394	0.99 (0.90-1.10)
≥60	115	138	0.84 (0.71-0.99)	239	327	0.73 (0.65-0.82)
Residence						
Rural	264	267	0.99 (0.88-1.12)	--	--	--
Urban	501	484	1.03 (0.95-1.13)	--	--	--
Highest education						
Primary school or lower	299	299	1.00 (0.89-1.12)	160	195	0.82 (0.71-0.95)
Middle school	226	212	1.07 (0.93-1.23)	522	503	1.04 (0.95-1.14)
High school or higher	239	240	1.00 (0.88-1.14)	740	711	1.04 (0.97-1.12)
BMI, kg/m ²						
<18.5	19	17	1.13 (0.71-1.94)	37	38	0.98 (0.72-1.39)
18.5-23.9	334	326	1.03 (0.92-1.15)	678	660	1.03 (0.95-1.11)
24.0-27.9	288	282	1.02 (0.91-1.15)	513	512	1.00 (0.92-1.10)
≥28	123	126	0.98 (0.82-1.18)	195	199	0.98 (0.85-1.13)
Height, cm						
<150.2	127	122	1.04 (0.87-1.26)	93	119	0.78 (0.65-0.94)
150.2-154.1	168	151	1.11 (0.95-1.31)	200	199	1.01 (0.88-1.16)
154.2-158.1	206	217	0.95 (0.83-1.09)	389	387	1.00 (0.91-1.11)
≥158.2	263	261	1.01 (0.89-1.14)	741	704	1.05 (0.98-1.13)
No. of first-degree relatives diagnosed with overall cancer						
0	598	573	1.04 (0.96-1.14)	1,000	1,023	0.98 (0.92-1.04)
1	136	144	0.95 (0.80-1.12)	340	329	1.03 (0.93-1.15)
≥2	30	34	0.89 (0.64-1.28)	83	57	1.45 (1.12-1.91)
No. of live birth						
Nulliparous	14	12	1.18 (0.68-2.29)	56	58	0.96 (0.75-1.27)
1	376	353	1.07 (0.96-1.19)	914	822	1.11 (1.04-1.19)
≥2	374	386	0.97 (0.88-1.07)	453	529	0.86 (0.79-0.93)
Age at menarche, years						
<12	60	60	1.00 (0.77-1.31)	127	105	1.21 (1.00-1.48)
13-14	255	266	0.96 (0.85-1.09)	599	567	1.06 (0.97-1.15)
15-16	276	275	1.00 (0.89-1.13)	519	553	0.94 (0.86-1.02)
≥17	173	150	1.16 (0.99-1.37)	178	184	0.97 (0.84-1.12)

Abbreviations: BMI, body mass index; PY, person-year; RR, relative risk; CI, confidence interval; E, expected number of cases; O, observed number of cases; --, not applicable.

^aAbsolute risk was recalibrated using the breast cancer incidence rates and non-breast cancer mortality rates in the CKB.

^bAbsolute risk was recalibrated using the breast cancer incidence rates and non-breast cancer mortality rates in Shanghai (Additional file 2).

Additional file 5. Performance of the breast cancer prediction model across different predicted risk cutoffs in the test subcohort of China Kadoorie Biobank

Cutoff of 10-year risk (%)	Percent of high-risk population (%) ^a	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)	Number needed to be screened to identify one case
0.4	84.74	94.90	15.34	0.82	99.75	121.51
0.5	70.89	86.17	29.23	0.89	99.65	111.96
0.6	56.17	74.27	43.96	0.97	99.57	102.92
0.7	42.93	65.05	57.24	1.11	99.55	89.81
0.8	32.19	55.58	67.99	1.27	99.52	78.80
0.9	24.27	45.63	75.88	1.38	99.47	72.39
1.0	18.26	38.11	81.88	1.53	99.44	65.22
1.1	13.61	31.80	86.52	1.72	99.42	58.27
1.2	10.32	26.21	89.80	1.87	99.40	53.57
1.3	7.78	21.60	92.32	2.04	99.38	49.01
1.4	5.73	17.23	94.36	2.21	99.35	45.23
1.5	4.13	12.38	95.93	2.20	99.33	45.45
1.6	3.05	9.95	97.00	2.40	99.32	41.68
1.7	2.18	6.80	97.85	2.29	99.30	43.75
1.8	1.64	5.10	98.39	2.29	99.29	43.71
1.9	1.13	3.88	98.89	2.52	99.29	39.75
2.0	0.79	2.67	99.22	2.48	99.28	40.36

^aHigh-risk population is defined as people whose 10-year predicted risk are higher than the corresponding cutoff.

Additional file 6. Performance of the breast cancer prediction model across different predicted risk cutoffs in the Shanghai Women's Health Study

Cutoff of 10-year risk (%)	Percent of high-risk population (%) ^a	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)	Number needed to be screened to identify one case
0.4	96.57	98.91	3.46	1.21	99.63	82.86
0.5	92.15	96.00	7.90	1.23	99.40	81.46
0.6	87.08	91.77	12.98	1.24	99.25	80.53
0.7	81.44	86.68	18.62	1.25	99.15	79.74
0.8	73.96	80.15	26.12	1.28	99.10	78.32
0.9	64.42	72.28	35.68	1.32	99.08	75.64
1.0	54.00	63.44	46.11	1.38	99.06	72.25
1.1	43.75	54.00	56.38	1.45	99.04	68.76
1.2	34.96	44.79	65.16	1.51	99.00	66.23
1.3	27.55	37.17	72.57	1.59	98.98	62.90
1.4	21.53	30.51	78.58	1.67	98.96	59.88
1.5	16.74	25.06	83.36	1.76	98.94	56.70
1.6	12.95	20.58	87.14	1.87	98.92	53.41
1.7	9.89	16.46	90.19	1.96	98.91	50.97
1.8	7.47	13.08	92.59	2.06	98.89	48.51
1.9	5.59	9.56	94.46	2.02	98.87	49.61
2.0	4.37	7.75	95.67	2.09	98.86	47.86

^aHigh-risk population is defined as people whose 10-year predicted risk are higher than the corresponding cutoff.

Additional file 7. Validation of the Asian America Breast Cancer Study model (AABCS) for predicting individual breast cancer risk in China Kadoorie Biobank (CKB) and Shanghai Women's Health Study (SWHS)

	CKB					SWHS				
	Overall AUC ^a	Adjusted AUC ^b	<i>E</i>	<i>O</i>	<i>E/O</i> (95% CI)	Overall AUC ^a	Adjusted AUC ^b	<i>E</i>	<i>O</i>	<i>E/O</i> (95% CI)
<i>Original AABCS model</i>										
Overall	0.574 (0.558-0.590)	0.543 (0.527-0.559)	4,325	2,287	1.89 (1.82-1.97)	0.553 (0.532-0.573)	0.556 (0.535-0.576)	1,639	1,409	1.16 (1.10-1.23)
1	--	--	180	154	1.17 (1.00-1.38)	--	--	102	117	0.87 (0.73-1.05)
2	--	--	250	181	1.38 (1.19-1.60)	--	--	132	125	1.06 (0.89-1.27)
3	--	--	276	162	1.70 (1.46-2.00)	--	--	143	132	1.08 (0.91-1.30)
4	--	--	318	212	1.50 (1.31-1.73)	--	--	140	117	1.19 (1.00-1.44)
5	--	--	354	181	1.96 (1.69-2.27)	--	--	147	121	1.22 (1.02-1.47)
6	--	--	363	193	1.88 (1.63-2.18)	--	--	161	132	1.22 (1.03-1.45)
7	--	--	406	269	1.51 (1.34-1.71)	--	--	176	131	1.35 (1.13-1.61)
8	--	--	465	292	1.59 (1.42-1.79)	--	--	181	163	1.11 (0.95-1.30)
9	--	--	599	306	1.96 (1.75-2.20)	--	--	200	169	1.18 (1.02-1.38)
10	--	--	1115	337	3.31 (2.97-3.69)	--	--	258	202	1.28 (1.11-1.47)
<i>Calibrated AABCS model^c</i>										
Overall	0.608 (0.0.592-0.623)	0.544 (0.529-0.590)	2,147	2,287	0.94 (0.90-0.98)	0.550 (0.530-0.570)	0.555 (0.535-0.575)	942	1,409	0.67 (0.63-0.71)
1	--	--	81	108	0.75 (0.62-0.91)	--	--	52	116	0.44 (0.37-0.54)
2	--	--	110	148	0.74 (0.63-0.88)	--	--	61	107	0.57 (0.47-0.70)
3	--	--	128	136	0.94 (0.80-1.13)	--	--	78	134	0.58 (0.49-0.69)
4	--	--	142	150	0.95 (0.81-1.12)	--	--	83	110	0.75 (0.63-0.92)
5	--	--	153	202	0.76 (0.66-0.87)	--	--	90	149	0.61 (0.52-0.72)
6	--	--	181	207	0.88 (0.76-1.01)	--	--	97	145	0.67 (0.57-0.79)

7	--	--	206	256	0.80 (0.71-0.91)	--	--	104	145	0.71 (0.61-0.85)
8	--	--	255	336	0.76 (0.68-0.85)	--	--	106	140	0.76 (0.64-0.90)
9	--	--	305	363	0.84 (0.76-0.93)	--	--	120	159	0.75 (0.64-0.88)
10	--	--	588	381	1.54 (1.39-1.71)	--	--	153	204	0.75 (0.65-0.86)

Abbreviations: AUC, area under the curve; E, expected number of cases; O, observed number of cases; CI, confidence interval; --, not applicable.

^aOverall AUC indicated the discriminating ability of the absolute risk predicted by our model.

^bAdjusted AUC were estimated adjusting for residence (urban/rural) and 10 age groups at entry, i.e. the prediction effect of age and residence was removed.

^cCalibrated using the breast cancer incidence rates and non-breast cancer mortality rates in China.