Supplementary Material

1: Modelling the conditional probability of the observed PRS given the total polygenotype in BOADICEA.

Definitions

The Kolmogorov definition for finite probabilities of the conditional probability of event A given event B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \qquad (s1.1)$$

Similarly, for continuous random variables we can define the conditional density of x given the occurrence of the value y of Y as

$$f_X(x|Y = y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}. \qquad (s1.2)$$

We can then write the conditional density of x given the occurrence of $y \in B$, where B represents a finite domain

$$f_X(x|y \in B) = \frac{\int_B f_{X,Y}(x, w)\, dw}{\int_B f_Y(z)\, dz}. \qquad (s1.3)$$

The finite probability of $x \in A$ is then

$$P(x \in A|y \in B) = \int_A f_X(w|y \in B)\, dw = \frac{\int_A \int_B f_{X,Y}(v, w)\, dw\, dv}{\int_B f_Y(z)\, dz}. \qquad (s1.4)$$

Implementation in BOADICEA

In BOADICEA we aimed at defining the finite conditional probability of the observed PRS (continuous variable) given the polygenotype (which is discrete under the Hypergeometric Polygenic Model (HPM)). We begin by first considering that both variables are continuous, and then discretised using equations (s1.3) and (s1.4).

The total polygene in BOADICEA is normally distributed in the general population[1] (see also main text equation (1)),

$$x_P \sim \mathcal{N}(0, 1). \qquad (s1.5)$$

Through genotyping of the known common breast cancer susceptibility variants (SNPs), we can observe part of x_P, the observed component summarised as the PRS. Therefore, we decomposed the polygenotype into an observed PRS component and an unobserved residual component

$$x_P = x_{PRS} + x_{R,} \qquad (s1.6)$$

where each component is independent and normally distributed such that

$$x_{PRS} \sim \mathcal{N}(0, \alpha^2), \tag{s1.7}$$

and

$$x_R \sim \mathcal{N}(0, 1-\alpha^2), \tag{s1.8}$$

where $0 \leq \alpha \leq 1$ is the proportion of the polygene captured by the PRS.

In order to calculate the conditional probability of x_{PRS} given x_P we need to define the densities of the total polygenotype, $f_{X_P}(x_P)$, and the joint distribution of the polygenotype and the PRS, $f_{X_{PRS},X_P}(x_{PRS}, x_P)$. As the polygenotype is normally distributed i.e.

$$f_{X_P}(x_P) = \frac{1}{\sqrt{2\pi}} e^{-x_P^2/2}. \tag{s1.9}$$

The PRS and residual components are independent and so their joint distribution is simply the product of their marginal distributions, so we have

$$f_{X_{PRS},X_R}(x_{PRS}, x_R) = \frac{1}{\alpha\sqrt{2\pi}} e^{-x_{PRS}^2/2\alpha^2} \frac{1}{\sqrt{2\pi(1-\alpha^2)}} e^{-x_R^2/2(1-\alpha^2)}. \tag{s1.10}$$

Then using equation (s1.6) we can eliminate x_R in favour of x_P to give the joint distribution of the polygenotype and the PRS

$$f_{X_{PRS},X_P}(x_{PRS}, x_P) = \frac{1}{\alpha\sqrt{2\pi}} e^{-x_{PRS}^2/2\alpha^2} \frac{1}{\sqrt{2\pi(1-\alpha^2)}} e^{-(x_P - x_{PRS})^2/2(1-\alpha^2)}. \tag{s1.11}$$

Substituting equations (s1.9) and (s1.11) into equation (s1.2) gives the conditional density

$$f_{X_{PRS}}(x_{PRS}|X_P = x_P) = \frac{1}{\sqrt{2\pi\,\alpha^2(1-\alpha^2)}} e^{-(x_{PRS} - x_P\,\alpha^2)^2/2\alpha^2(1-\alpha^2)}, \tag{s1.12}$$

which also corresponds to equation (4) in MacInnis et al.[2]

Discrete Case – formulation under the HPM

Under the HPM, the polygenotype is discretised using a binomial distribution into $2N+1$ bins, with bin intervals given by

$$B(k; 2N) = \left[NQF\left(BCDF\left(k-1, 2N, \frac{1}{2}\right), 1\right), NQF\left(BCDF\left(k, 2N, \frac{1}{2}\right), 1\right)\right], \tag{s1.13}$$

where $k = 0, \ldots, 2N$ indexes the bins, $BCDF$ is the binomial cumulative distribution function and NQF is the normal quantile function. Then using equations (s1.9) and (s1.11) in equation (s1.3), the conditional density of observing x_{PRS} given $x_P \in B(k; n)$ is

$$f_{X_{PRS}}\left(x_{PRS} \middle| x_P \in B(k; 2N)\right) = \frac{\int_{B(k;2N)} f_{X_{PRS},X_P}(x_{PRS}, w)\, dw}{\int_{B(k;2N)} f_{X_P}(z)\, dz}, \tag{s1.14}$$

where the denominator is just the binomial probability.

Since $x_{PRS}$is assumed to be continuous, equation (s1.14) describes a conditional density (rather than a probability), and likelihoods calculated with it will be likelihood densities. However, as x_{PRS} is the same for all terms in the likelihood, and the measure is independent of $B(k;n)$, the ratios of such likelihood densities are finite probabilities. Pathogenic variant carrier probabilities and the predicted cancer risks in BOADICEA can be calculated as the ratios of such likelihood densities, when a measured PRS is present.

In the limit $\alpha \to 1$ (where the total polygene is completely observed), the second term in equation (s1.11) becomes a Dirac delta function at x_{PRS}, so when we perform the integration in equation (s1.3) the density for each bin is 0 except for the bin that contains $x_P = x_{PRS}$. This corresponds with Lange's elimination of incompatible genotypes[3].

Similarly, in the limit $\alpha \to 0$ (where the polygene is completely unobserved), the first term in equation (s1.11) becomes a Dirac delta function at $x_{PRS} = 0$, and the density in equation (s1.14) becomes the same for all bins.

As such, the use of a conditional probability to include the effects of the PRS in the penetrance function, equation (5) in the main text, can be thought of as an extension of Lange's removal of incompatible genotypes to the case of a partial observation of a continuous genotype (in Lange's case the probability of making the observation given the dummy is a Kronecker delta function). We then demonstrate how this is discretised within the HPM.

The approach presented here to include the explicit joint effects of common genetic variants has the advantage that it can flexibly adapt to a different PRS by varying the value of α to reflect the proportion of the polygenotype explained by the specific genotyped variants in that PRS. In the current implementation, only the woman for whom a breast cancer risk is calculated, is allowed to have an observed PRS, although in principle PRS information on any/multiple pedigree members could be used.

2: Incorporating Risk Factors and Mammographic Density (BI-RADS) in BOADICEA

Distributions and Relative Risks

In BOADICEA we assume population distributions to be static with age, but the relative risks (RRs) are allowed to vary with age. In principle, the RR and risk factor distributions could also be cohort dependent but in this version we have assumed that they do no vary by birth cohort. The set of risk factors included were based on the synthetic model of Garcia-Closas et al.[4], but the RR estimates were updated where necessary considering data from large well-designed studies and population distributions obtained from surveillance data summarised in Pal Choudhury et al.[5]. We have assumed that population distributions are those given in Pal Choudhury et al.[5] for women under age 50. Tables s2.1-s2.10 summarise the RR estimates and population risk factor distributions, and the sources, for each risk factor used in the model.

Mammographic density

Mammographic density	Category	Population distribution	Relative risk up to age 20	Relative risk for age 20 to 49	Relative risk for age 50 and over
BI-RADS a Fatty	1	0.041	1.0	0.41	1.00
BI-RADS b Scattered Density	2	0.352	1.0	1.00	2.04
BI-RADS c Heterogeneously Dense	3	0.470	1.0	1.75	2.81
BI-RADS d Extremely Dense	4	0.137	1.0	3.33	4.08
Source		Tice et al.[6] Table 2 (Whites <50 years old)		Nelson et al.[7] Table 2	McCormack and dos Santos Silva[8] Figure 2

Table s2.1: Mammographic density population distribution and relative risk estimates used in BOADICEA.

Age at menarche

Age at menarche	Category	Population distribution	Relative risk up to age 20	Relative risk for age 20 and over
< 11	1	0.122	1.0	1.19
11	2	0.216	1.0	1.09
12	3	0.268	1.0	1.07
13	4	0.215	1.0	1.00
14	5	0.115	1.0	0.98
15	6	0.044	1.0	0.92
>15	7	0.019	1.0	0.82
Source		Pal Choudhury et al.[5] using data from UK Generations Study[9]		Collaborative Group on Hormonal Factors in Breast Cancer[10] Figure 3 A

Table s2.2: Age at menarche population distribution and relative risk estimates used in BOADICEA.

Age at menopause

Age at menopause	Category	Population distribution	Split age	Relative risk up to split age	Relative risk for split age and above
< 40	1	0.022	39	1.0	0.67
40-44	2	0.070	40	1.0	0.73
45-49	3	0.234	45	1.0	0.86
50-54	4	0.533	50	1.0	1.00
> 54	5	0.14	55	1.0	1.12
Source		Pal Choudhury etal.[5] using data from UK Generations Study[9]			Collaborative Group on Hormonal Factors in Breast Cancer[10] Figure 3 B

Table s2.3: Age at menopause population distribution and relative risk estimates used in BOADICEA.

Parity

Parity	Category	Population distribution	Relative risk up to age 20	Relative risk for age 20 and over
Nulliparous	1	0.28	1.0	1.00
1 birth	2	0.16	1.0	0.87
2 births	3	0.33	1.0	0.81
> 2 births	4	0.23	1.0	0.71
Source		Pal Choudhury et al.[5] using data from Cohort Fertility Tables[11] Table 3		Reeves et al.[12] Figure 2

Table s2.4: Parity population distribution and relative risk estimates used in BOADICEA.

Age at First Live Birth

Age at first live birth was considered in age groups. For each age group we defined a "split" age. Below this age we assumed that the relative risk is 1.0, and at the split age or above we assumed the

published RR estimate. Note that the population distribution does not sum to 1.0, but rather to 0.72, the proportion of women in the population who are parous.

Age at first live birth	Category	Population distribution	Parous distribution	Split age	Relative risk up to split age	Relative risk for split age and above
< 20	1	0.12	0.1667	19	1.0	1.00
20-24	2	0.23	0.3194	20	1.0	1.01
25-29	3	0.23	0.3194	25	1.0	1.11
>29	4	0.14	0.1944	30	1.0	1.24
Source		Pal Choudhury et al.[5] using data from Cohort FertilityTables[11] Table 2				Reeves et al.[12] Figure 2

Table s2.5: Age at first live birth population distribution and relative risk estimates used in BOADICEA.

Oral Contraceptive (OC) Use

Current use of OC	Category	Population distribution	Relative risk up to current age	Relative risk from current age up to 50	Relative risk for age 50 and over
Never	1	0.185	1.0	1.00	1.0
Former	2	0.684	1.0	1.12	1.12
Current	3	0.131	1.0	1.33	1.12
Source		Pal Choudhury et al.[5] using data from Health Survey for England 2005[13] and 2006[14]		Hunter[15] Table 2	

Table s2.6: Current use of oral contraception population distribution and relative risk estimates used in BOADICEA.

Menopausal Hormone Replacement Therapy (HRT) Use

Current use of HRT	Category	Population distribution	Relative risk up to current age	Relative risk for current age to current age + 5	Relative risk for current age + 5 and over
Never/Former any type	1	0.913	1.0	1.00	1.0
Current E-type	2	0.010962	1.0	1.57	1.0
Current other/unknown type	3	0.076038	1.0	2.19	1.0
Source		Pal Choudhury et al.[5] using data from Health Survey for England 2005[13] and 2006[14], Parkin[16]		Pal Choudhury et al.[5] using data from Beral et al.[17] Figure 5	

Table s2.7: Use of menopause hormone replacement therapy population distribution and relative risk estimates used in BOADICEA.

Body Mass Index (BMI)

BMI	Category	Population distribution	Relative risk up to age 20	Relative risk for age 20 to 49	Relative risk for age 50 and over
<18.5	1	0.033	1.0	1.28	1.00
18.5-24.9	2	0.357	1.0	1.00	1.00
25-29.9	3	0.358	1.0	0.92	1.13
>= 30	4	0.252	1.0	0.74	1.25
Source		Pal Choudhury et al.[5] using data from Health Survey for England 2005[13] and 2006[14]		Nelson et al.[7] Table 1	Pal Choudhury et al.[5] using data from Beral et al.[17] Figure 5

Table s2.8: Body Mass Index (BMI) population distribution and relative risk estimates used in BOADICEA.

Height

Height (cm)	Category	Population distribution	Relative risk up to age 20	Relative risk age 20 and over
≤152.90	1	0.0625	1.0	0.82
152.91-159.64	2	0.25	1.0	0.91
159.65-165.95	3	0.375	1.0	1.00
165.96-172.69	4	0.25	1.0	1.10
≥172.70	5	0.0625	1.0	1.22
Source		Assuming normally distributed with mean 162.81cm and standard deviation 6.452cm Pal Choudhury et al.[5] using data from Health Survey for England 2005[13] and 2006[14]		RR of 1.17 per 10cm Zhang et al.[18] Green et al.[19]

Table s2.9: Height population distribution and relative risk estimates used in BOADICEA.

Alcohol Intake

Alcohol Intake per day	Category	Population distribution	Relative risk up to age 20	Relative risk for age 20 and over
0g	1	0.106	1.0	1.00
<5g	2	0.307	1.0	1.01
5-14g	3	0.276	1.0	1.03
15-24g	4	0.133	1.0	1.13
25-34g	5	0.080	1.0	1.21
35-44g	6	0.036	1.0	1.32
>=45g	7	0.061	1.0	1.46
Source		Pal Choudhury et al.[5] using data from Health Survey for England 2002[20]		Hamajima et al.[21] Table 2

Table s2.10: Daily alcohol intake in grams population distribution and relative risk estimates used in BOADICEA.

In principle our modelling approach, equation (2) of the main text, allows the relative risks to depend on the major genotypes considered in the model (*BRCA1*, *BRCA2*, *PALB2*, *CHEK2* and *ATM*), however since the risk factor effects are still not well established for pathogenic variant carriers, in this implementation we assumed that the effects of the risk factors are the same for all pathogenic variant carriers and non-carriers. Further, we assume that the risk factors are independent of each other and of mammographic density, ignoring interactions between them.

The time required to calculate the age-specific baseline incidence in equation (2) of the main text is proportional to the product of the number of categories for each individual risk factor. For the full set of risk factors described above, this would lead to lengthy runtimes, which would compromise

real-time risk calculations in clinical practice. In order to reduce the runtime of the code we combined a number of factors into a single factor with a smaller number of categories. This was achieved by first calculating the joint distribution of multiple risk factors and then reducing the overall number of categories, by amalgamating adjacent categories. The proband's risk factor category was not amalgamated with any other categories, but assigned a category of its own. The RR associated with each new category was taken as the weighted average of its constituent categories, with a probability mass given by their sum. To ensure that individuals remain in the same categories for their entire life, only factors whose RRs are independent of age were combined. For this purpose, we combined age at menarche, parity, height and alcohol intake into a single factor. This approximation had only a minimal impact on BOADICEA risk predictions (at the 4^{th} significant figure), and was therefore considered negligible, while achieving a significant (160 fold) speed up.

3: Age-Specific CHEK2 Relative Risks

Recently, the largest study of breast cancer risks for *CHEK2* 1100delC pathogenic variant carriers has been published[22], which provided age-specific relative risk estimates. In this analysis, the best fitting model for the relative risk (compared to population risk) depends on age according to the following formula:

$$RR(t) = \begin{cases} 1.0 & \text{if } t < 20, \\ 3.190649 \times \exp\left(-0.0148367 \times (\mathrm{t} - 30)\right) & \text{if } t \geq 20, \end{cases} \qquad (s3.1)$$

where t is age. These age specific RR's have been incorporated into BOADICEA.

Figure s3.1 demonstrates the effects of including the updated *CHEK2* relative risk estimates in comparison to the previous BOADICEA model[23] on breast cancer risk. The lifetime risk of developing breast cancer for a *CHEK2* 1100delC pathogenic variant carrier is 22.1% by age 80.

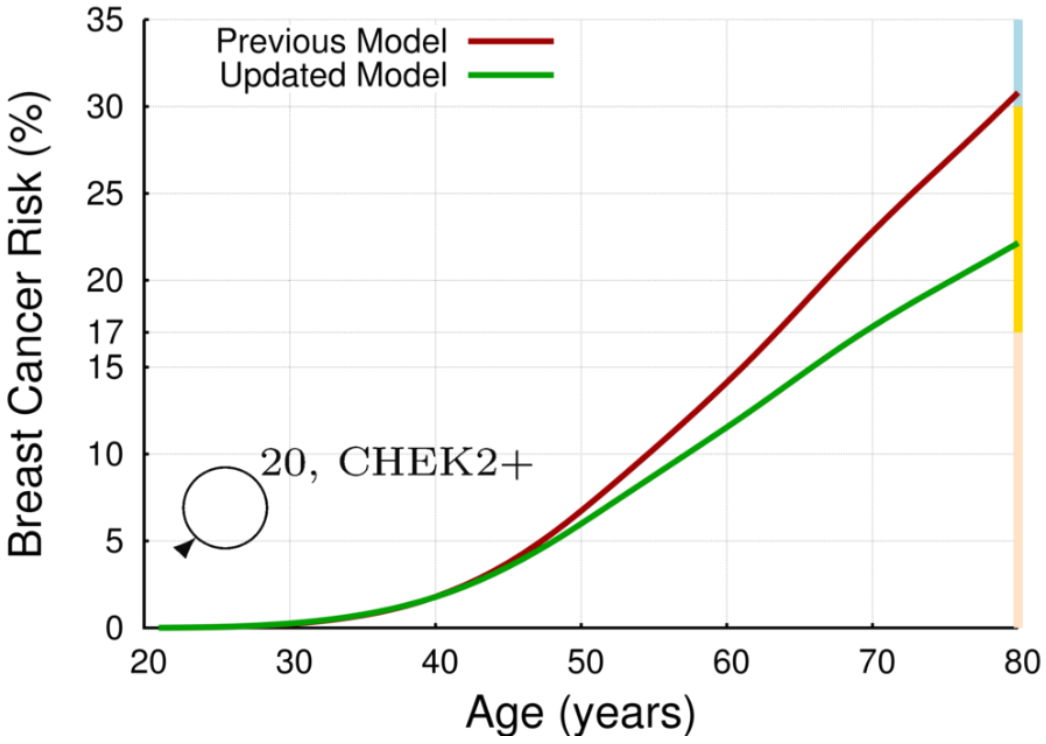


Figure s3.1: This shows the BOADICEA breast cancer risk for 20 year old female up to age 80, born in 1985, comparing the previous and current models. The right hand y-axis of the graph is shaded to indicate the familial breast cancer risk categories based on the NICE guidelines[24]: 1) near population risk shaded in pink ($< 17\%$), 2) moderate risk shaded in yellow ($\geq 17\%$ and $< 30\%$) and 3) high risk, shaded in blue ($\geq 30\%$). Predictions based on UK breast cancer incidences.

4: Inclusion of Spanish Incidences

BOADICEA uses birth-year and geographic-region cohort specific incidences. Here the model was further extended to allow for the option to use incidences from Spain using the methods described in[25].

The cancer incidences were taken from two sources:

- Incidences for the period 1968-2012 were taken from Cancer Incidence in Five Continents (CI5)[26-34].
- Incidence from 2015 were supplied by the Spanish Network of Cancer Registries (REDECAN)[35].

Data form CI5 came from 15 regions of Spain, which do not cover the entire country, and not all regions span the time period 1968 to 2012, summarised in Table s4.1. As regions have different population sizes, we created the national incidences using weighted averages (based on the age specific numbers of cases).

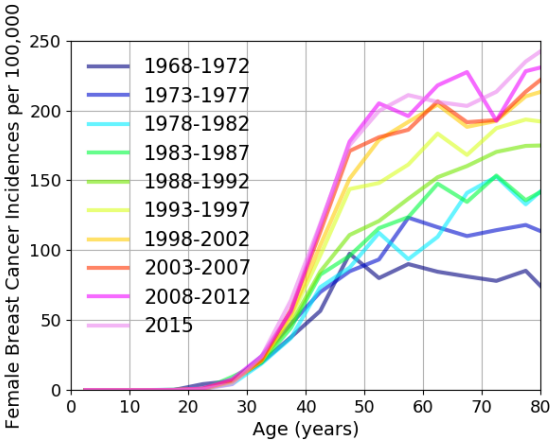
Region	1968-1972	1973-1977	1978-1982	1983-1987	1988-1992	1993-1997	1998-2002	2003-2007	2008-2012
Zaragoza	✓	✓	✓	✓	✓	✓	✓	×	×
Navarra	×	✓	✓	✓	✓	✓	✓	✓	✓
Tarragona	×	×	✓	✓	✓	✓	✓	✓	✓
Murcia	×	×	×	✓	✓	✓	✓	✓	✓
Granada	×	×	×	✓	✓	✓	✓	✓	✓
Basque Country	×	×	×	✓	✓	×	✓	✓	✓
Asturias	×	×	×	×	✓	✓	✓	✓	✓
Mallorca	×	×	×	×	✓	✓	×	✓	✓
Albacete	×	×	×	×	✓	✓	✓	✓	✓
Canary Islands	×	×	×	×	×	✓	✓	✓	✓
Cuenca	×	×	×	×	×	✓	✓	✓	✓
Girona	×	×	×	×	×	✓	✓	✓	✓
La Rioja	×	×	×	×	×	×	×	✓	✓
Ciudad Real	×	×	×	×	×	×	×	✓	✓
Castellon	×	×	×	×	×	×	×	×	✓

Table s4.1: Regions covered by data from Cancer Incidences in Five Continents for each calendar period.

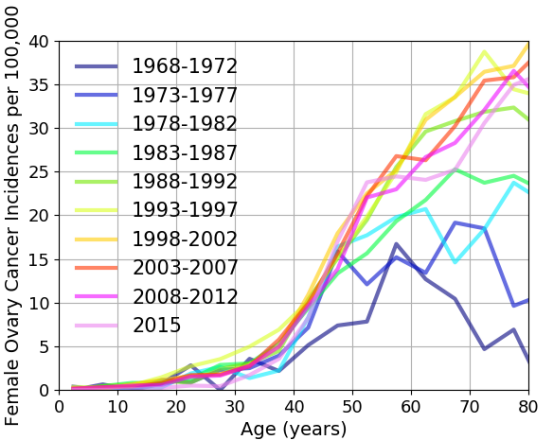
National incidences for 2015 were supplied by REDECAN. This data does not include incidences on male breast cancer, so the latest available data from CI5 was used instead.

Figure s4.1 plots the national age-specific incidences for each caner site for each calendar period for. Figures s4.2-s4.7 compare the Spanish incidences to those from other countries (UK, USA and Sweden). In general, female breast cancer incidences for Spain are lower than those of the other three countries.

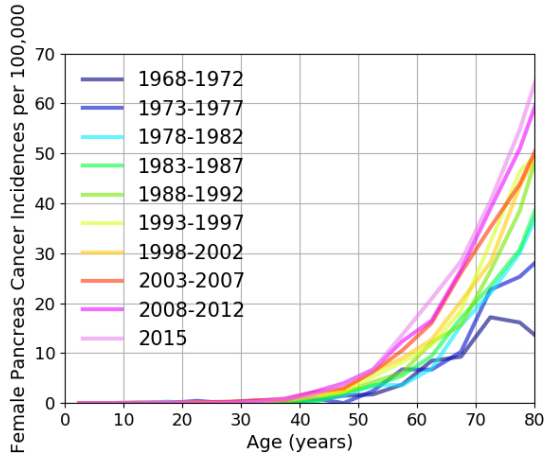
These calendar specific incidences were then used to create smoothed birth cohort specific incidences for use in BOADICEA, as previously described[25].



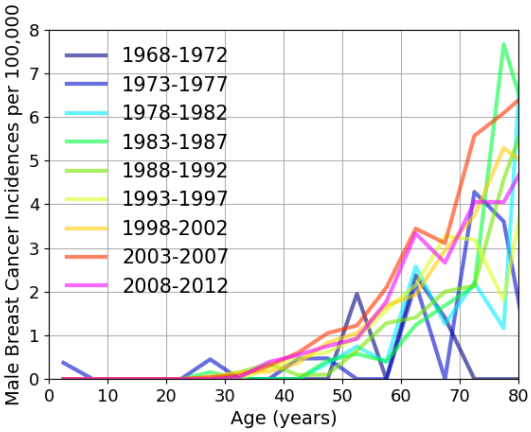
(a) Spanish female breast cancer incidence by year.



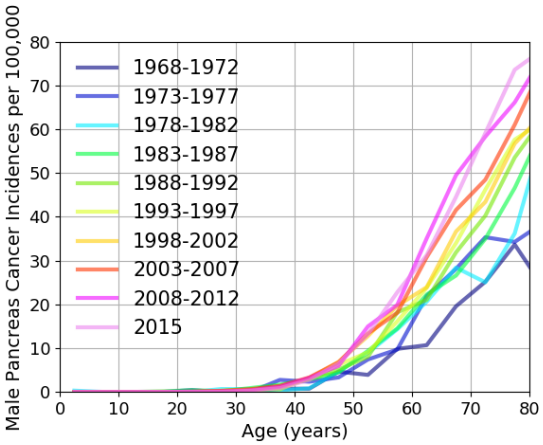
(b) Spanish female ovarian cancer incidence by year.



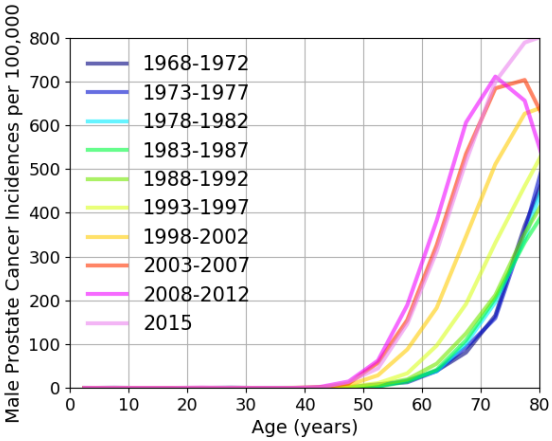
(c) Spanish female pancreatic cancer incidence by year.



(d) Spanish male breast cancer incidence by year.

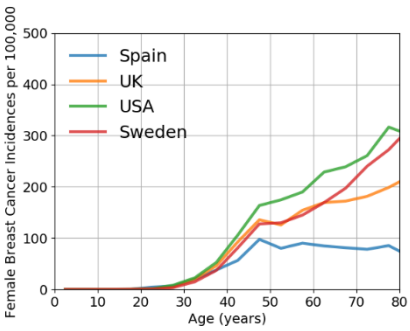


(e) Spanish male pancreatic cancer incidence by year.

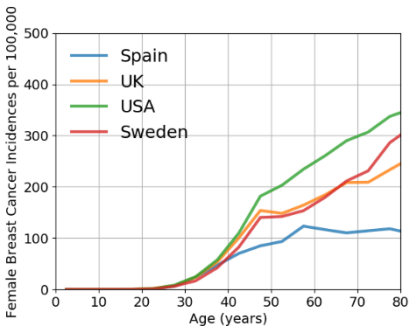


(f) Spanish male prostate cancer incidence by year.

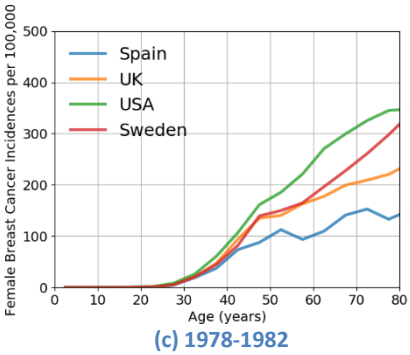
Figure s4.1: Comparison of incidences by calendar period.



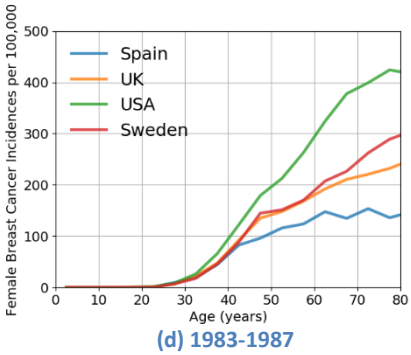
(a) 1968-1972



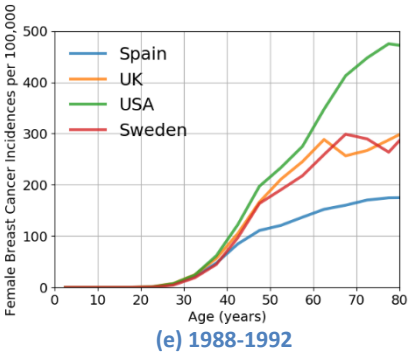
(b) 1973-1977



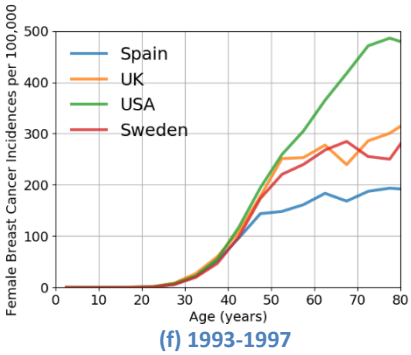
(c) 1978-1982



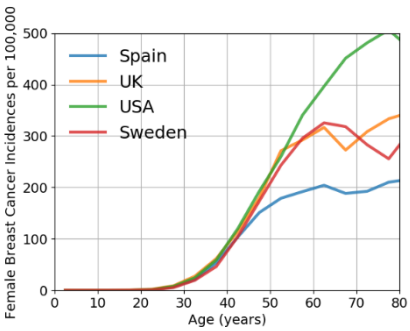
(d) 1983-1987



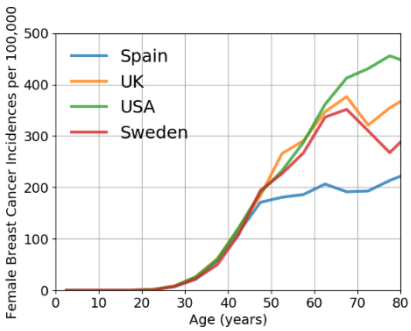
(e) 1988-1992



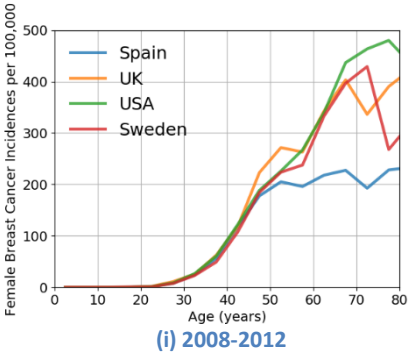
(f) 1993-1997



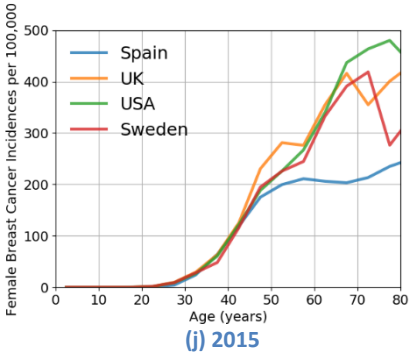
(g) 1998-2002



(h) 2003-2007

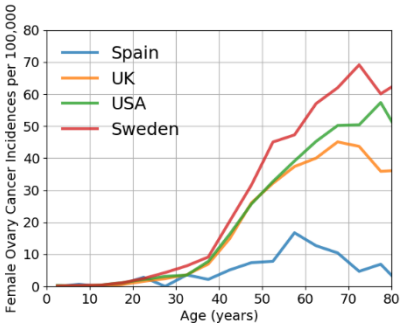


(i) 2008-2012

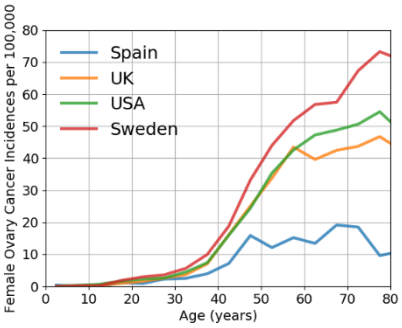


(j) 2015

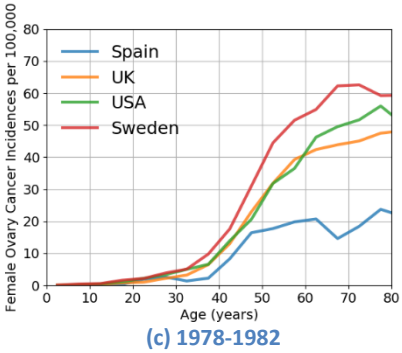
Figure s4.2: Comparison of female breast cancer incidences by country.



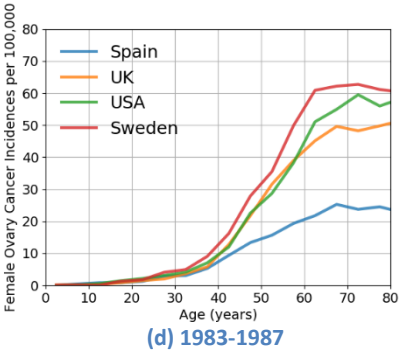
(a) 1968-1972



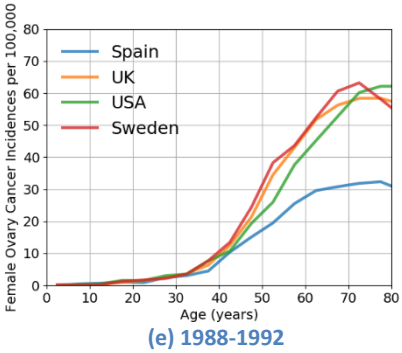
(b) 1973-1977



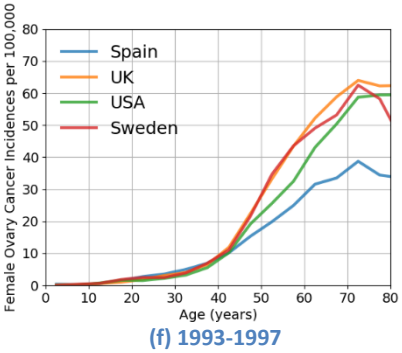
(c) 1978-1982



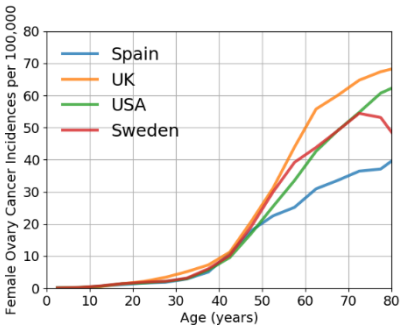
(d) 1983-1987



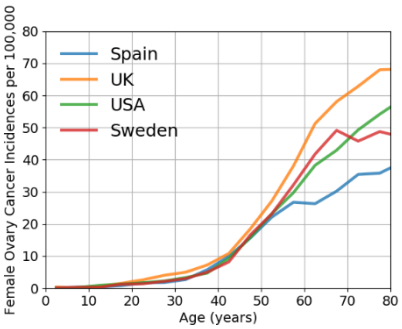
(e) 1988-1992



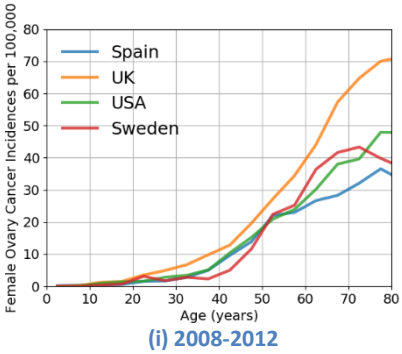
(f) 1993-1997



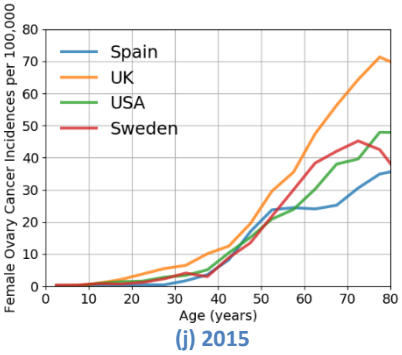
(g) 1998-2002



(h) 2003-2007

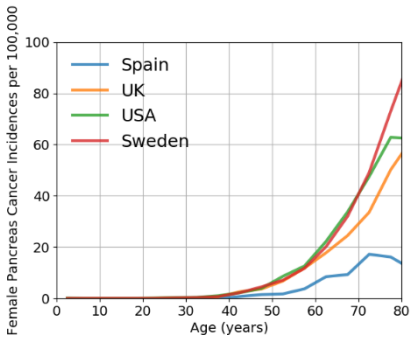


(i) 2008-2012

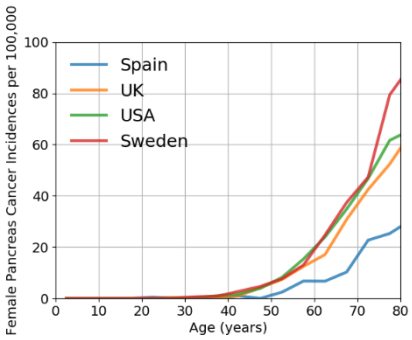


(j) 2015

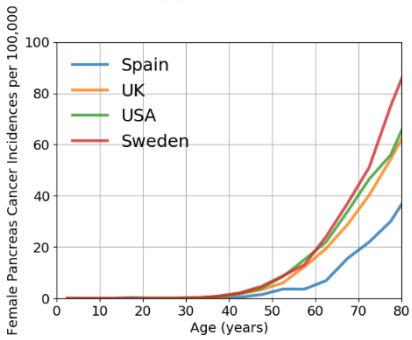
Figure s4.3: Comparison of female ovarian cancer incidences by country.



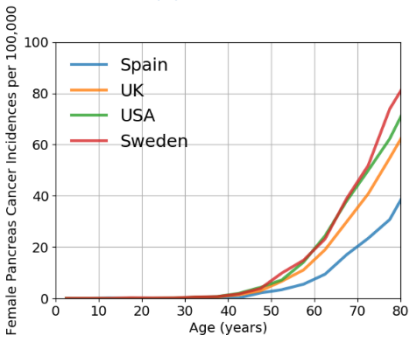
(a) 1968-1972



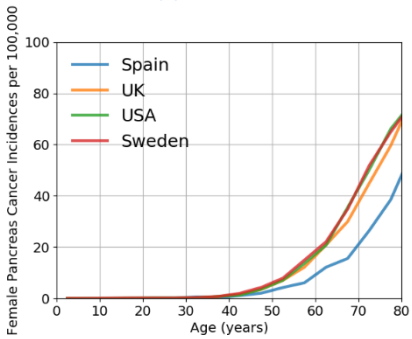
(b) 1973-1977



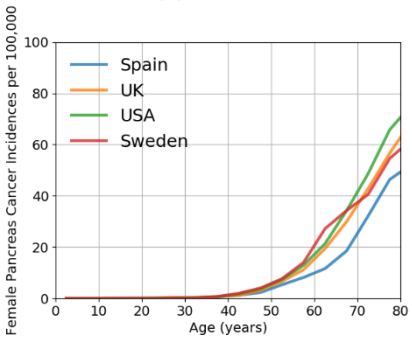
(c) 1978-1982



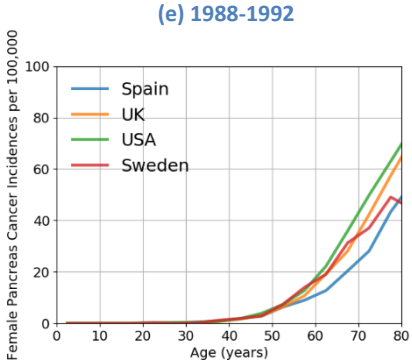
(d) 1983-1987



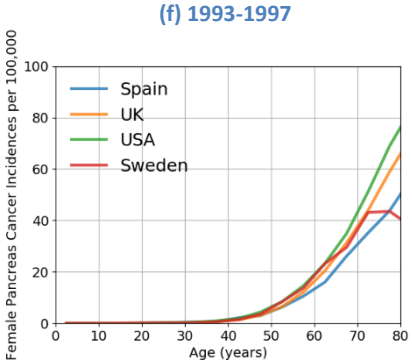
(e) 1988-1992



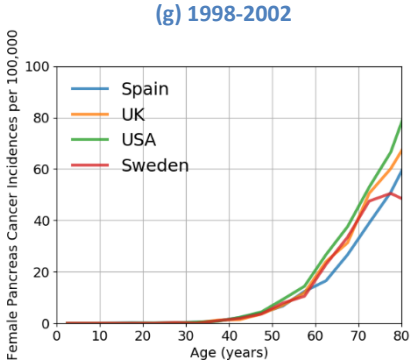
(f) 1993-1997



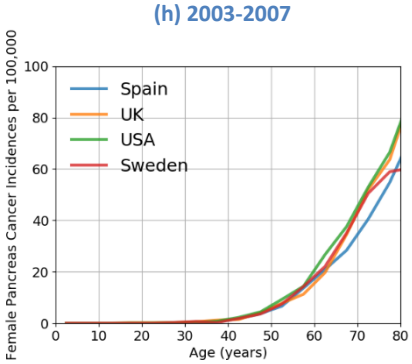
(g) 1998-2002



(h) 2003-2007

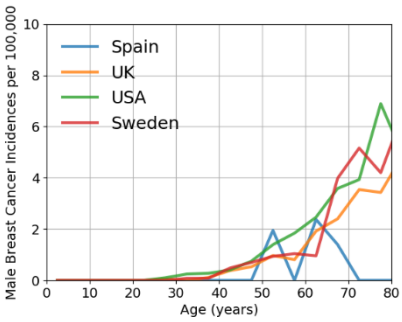


(i) 2008-2012

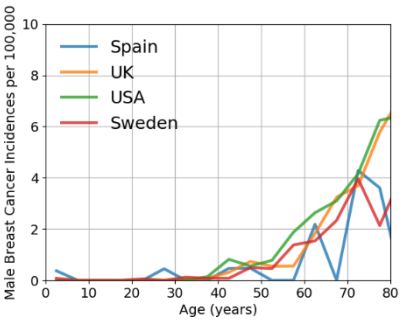


(j) 2015

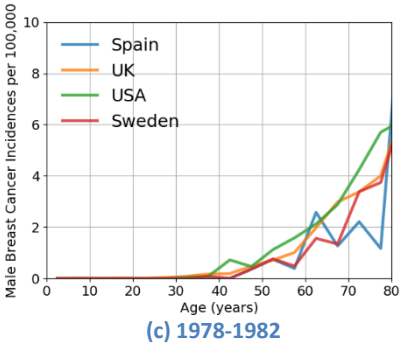
Figure s4.4: Comparison of female pancreatic cancer incidences by country.



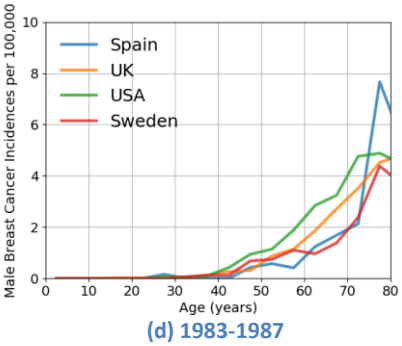
(a) 1968-1972



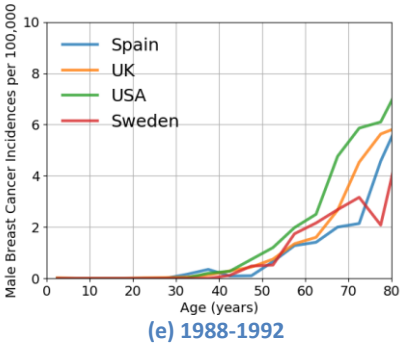
(b) 1973-1977



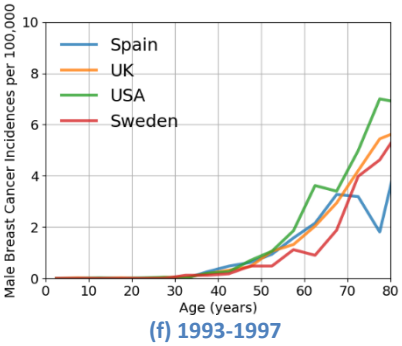
(c) 1978-1982



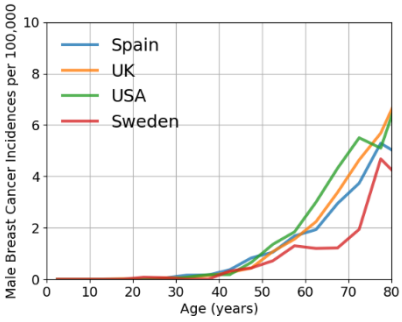
(d) 1983-1987



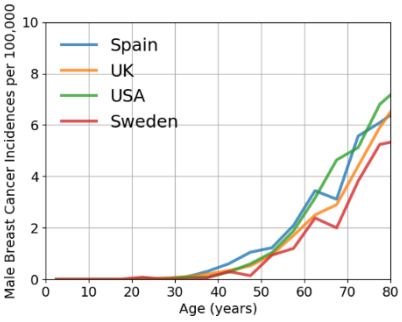
(e) 1988-1992



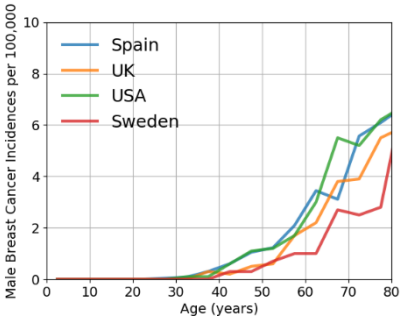
(f) 1993-1997



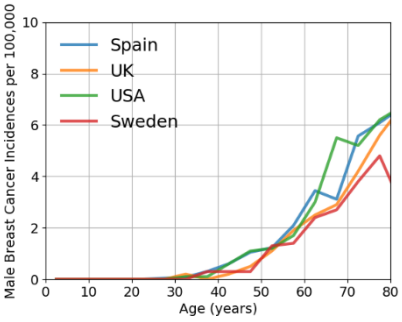
(g) 1998-2002



(h) 2003-2007

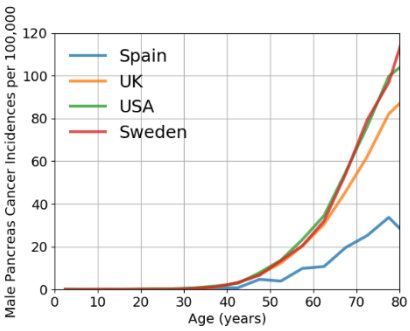


(i) 2008-2012

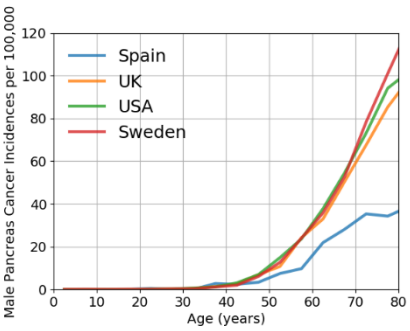


(j) 2015

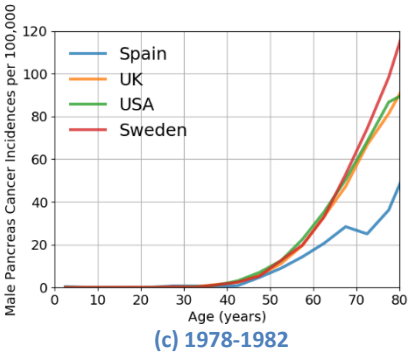
Figure s4.5: Comparison of male breast cancer incidences by country.



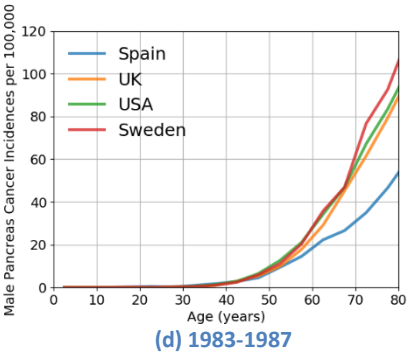
(a) 1968-1972



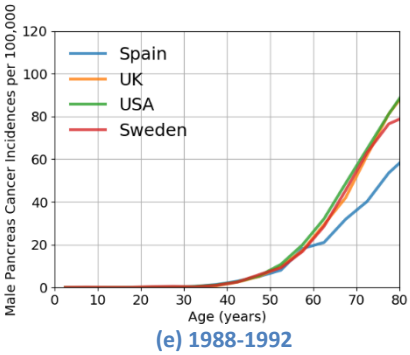
(b) 1973-1977



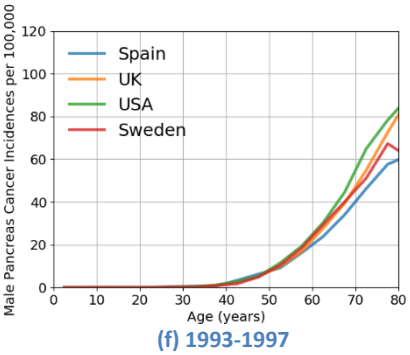
(c) 1978-1982



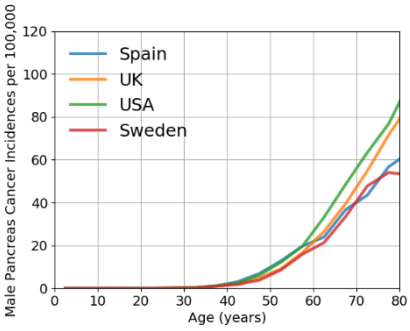
(d) 1983-1987



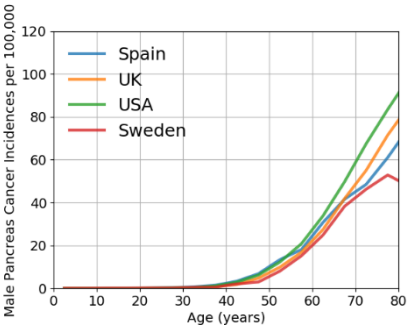
(e) 1988-1992



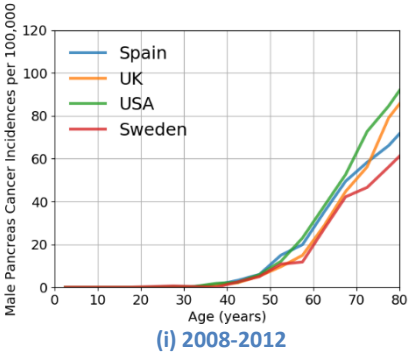
(f) 1993-1997



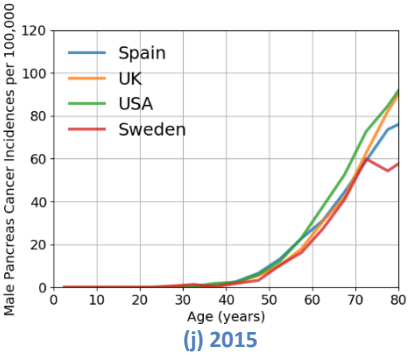
(g) 1998-2002



(h) 2003-2007

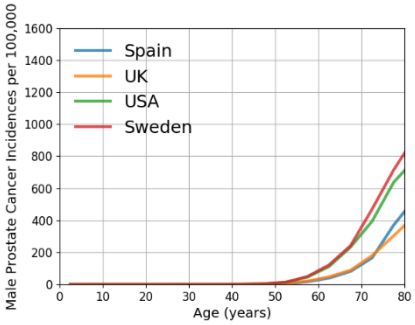


(i) 2008-2012

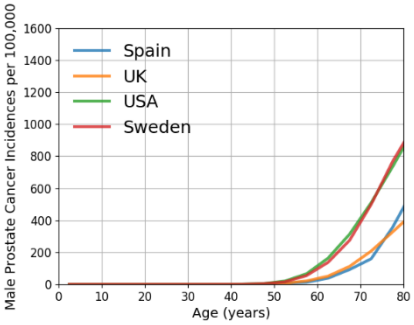


(j) 2015

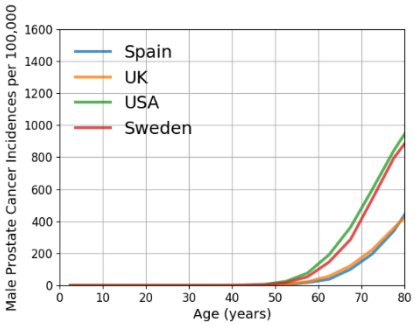
Figure s4.6: Comparison of male pancreatic cancer incidences by country.



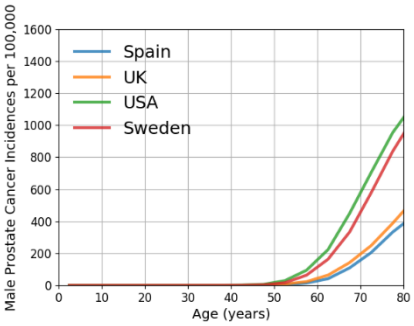
(a) 1968-1972



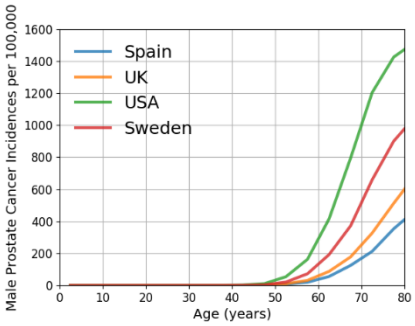
(b) 1973-1977



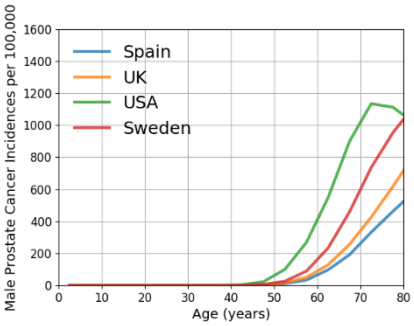
(c) 1978-1982



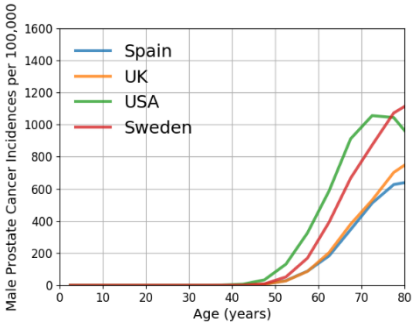
(d) 1983-1987



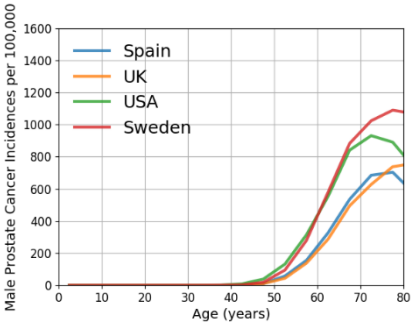
(e) 1988-1992



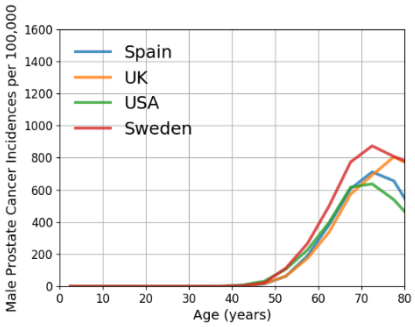
(f) 1993-1997



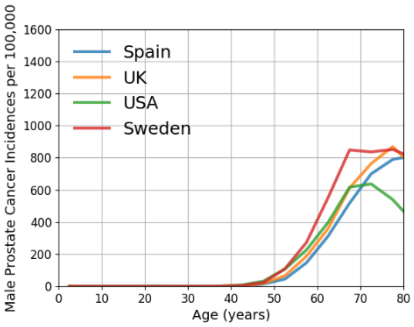
(g) 1998-2002



(h) 2003-2007



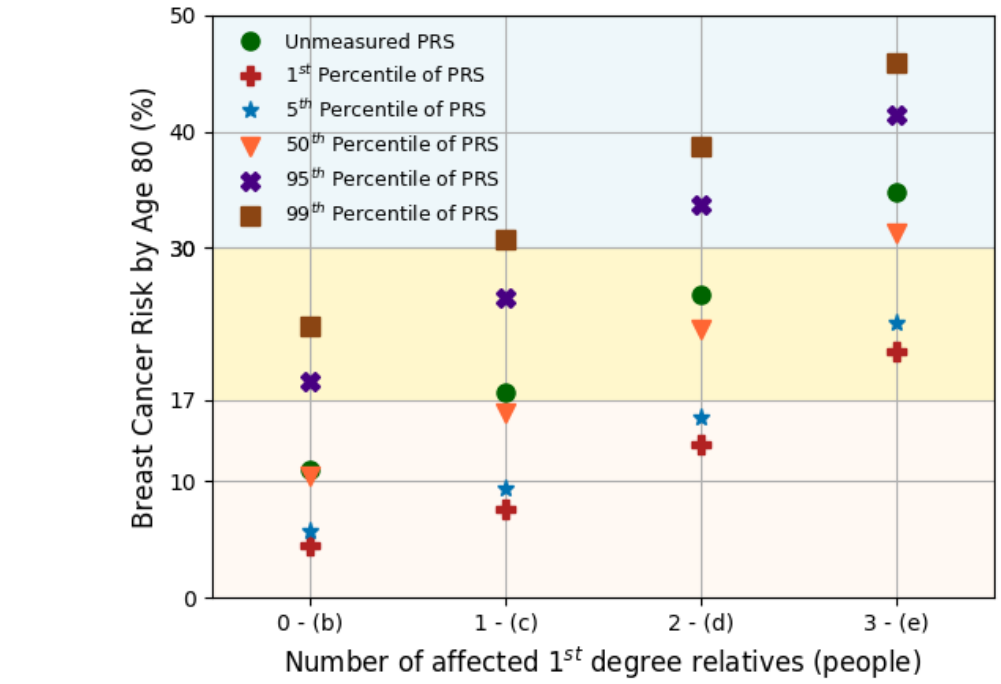
(i) 2008-2012



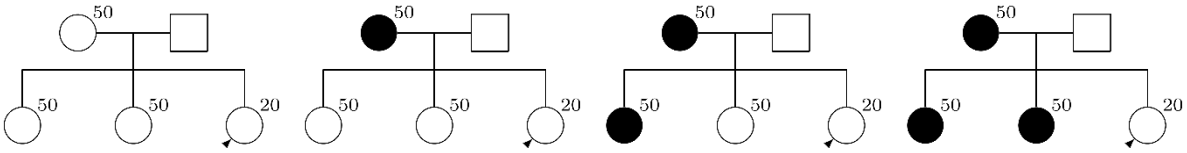
(j) 2015

Figure s4.7: Comparison of male prostate cancer incidences by country.

5: Family History and Polygenic Risk Scores



(a)



(b) (c) (d) (e)

Figure s5.1: Figure (a)) shows the lifetime risk predicted by BOADICEA on the basis of Family History (FH) and Polygenic Risk Score (PRS). Figures (b)-(e) show the assumed FH's considered on the x-axis of figure (a). The backgrounds of figure (a) is shaded to indicate the familial breast cancer risk categories based on the NICE guidelines[24]: 1) near population risk shaded in pink ($< 17\%$), 2) moderate risk shaded in yellow ($\geq 17\%$ and $< 30\%$) and 3) high risk, shaded in blue ($\geq 30\%$).

Family History	Absolute difference in risk between those at 1% and 99% of PRS	Variance of Log(risk)	Percentile with risk equal to risk with unmeasured PRS
(b)	18.8 (%)	0.133	55.9
(c)	23.2 (%)	0.095	63.6
(d)	25.6 (%)	0.059	69.2
(e)	24.7 (%)	0.031	72.9

Table s5.1: Joint effects of FH and PRS on predicted BC risk.

6: Risk Probability Trees

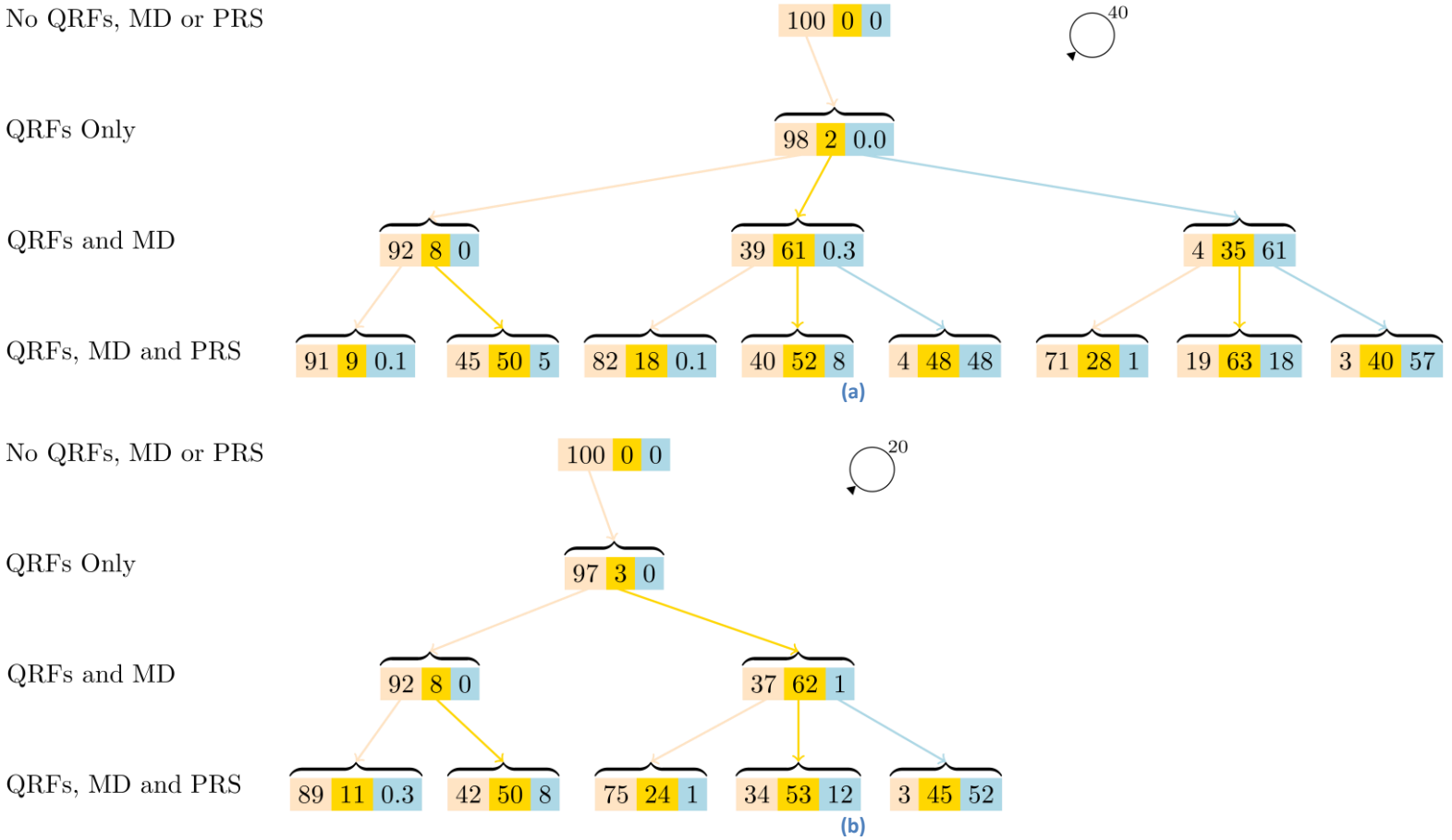


Figure s6.1. Probability Trees for (a) 10-year and (b) lifetime risks, for a woman with unknown family history, based on BOADICEA. Starting at the top of each tree, the figures show the percent of women reclassified by adding in more information to the breast cancer risk prediction, as indicated by the captions on the left (questionnaire-based risk factors (QRFs), mammographic density (MD) and Polygenic Risk Score (PRS)). Note that in comparison to figure s6.2, here MD is added in before the PRS. Each triplet of numbers is the percentage of women who fall into the familial breast cancer risk categories based on the NICE guidelines[24]: 1) near population risk shaded in pink ($< 17\%$ for lifetime risk and $< 3\%$ for 10-year risk), 2) moderate risk shaded in yellow ($\geq 17\%$ and $< 30\%$ for lifetime risk and $\geq 3\%$ and $< 8\%$ for 10-year risk) and 3) high risk, shaded in blue ($\geq 30\%$ for lifetime risk and $\geq 8\%$ for 10-year risk). Percentages identically equal to zero are denoted as "0" (i.e. no women fall into that category), while percentages less than 0.1 are denoted by "0.0" (i.e. a very small number of women fall into that category).

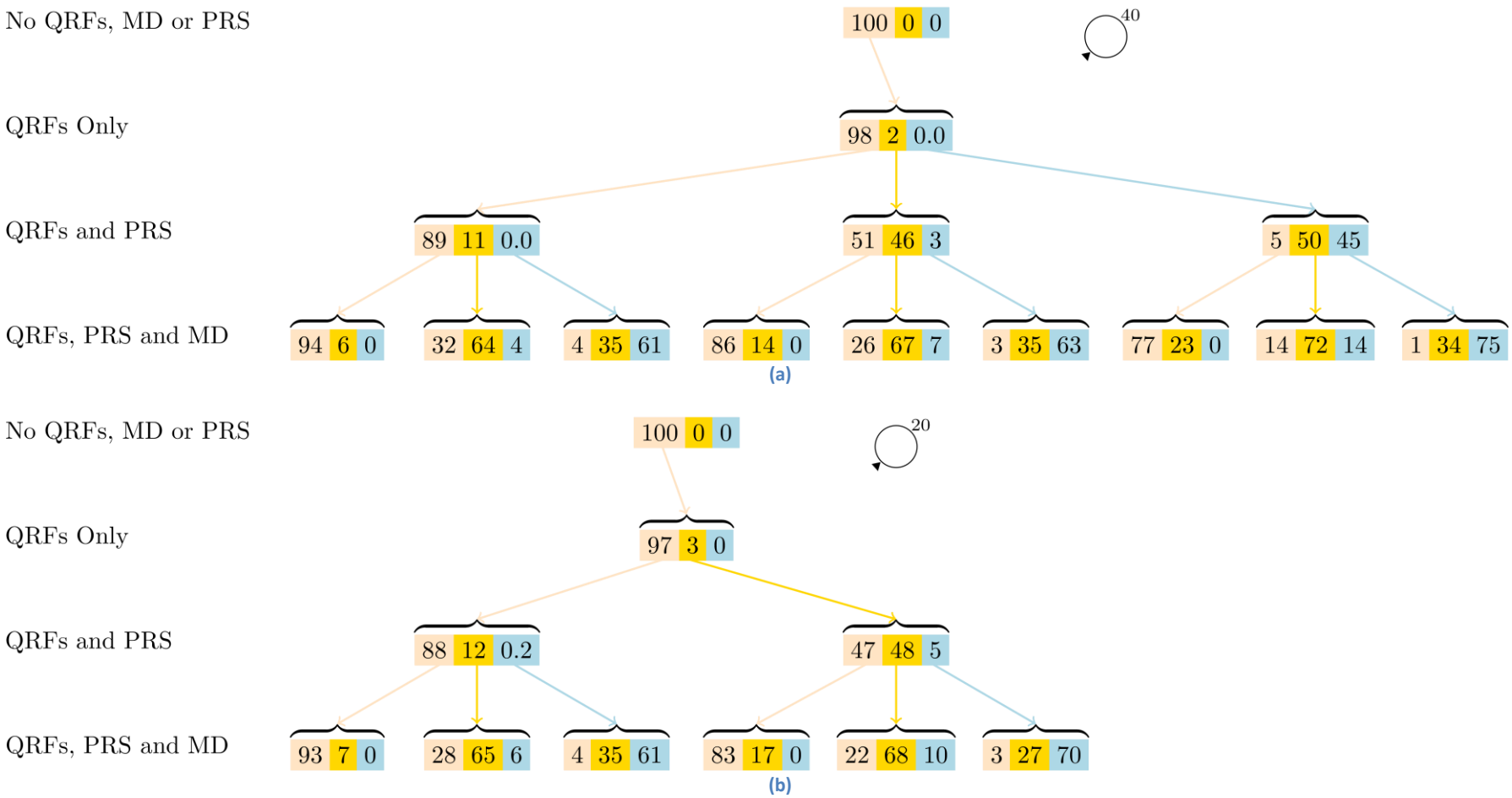


Figure s6.2. Probability Trees for (a) 10-year and (b) lifetime risks, for a woman with unknown family history, based on BOADICEA. Starting at the top of each tree, the figures show the percent of women reclassified by adding in more information to the breast cancer risk prediction, as indicated by the captions on the left (questionnaire-based risk factors (QRFs), mammographic density (MD) and Polygenic Risk Score (PRS)). Note that in comparison to figure s6.1, here the PRS is added in before MD. Each triplet of numbers is the percentage of women who fall into the familial breast cancer risk categories based on the NICE guidelines[24]: 1) near population risk shaded in pink ($< 17\%$ for lifetime risk and $< 3\%$ for 10-year risk), 2) moderate risk shaded in yellow ($\geq 17\%$ and $< 30\%$ for lifetime risk and $\geq 3\%$ and $< 8\%$ for 10-year risk) and 3) high risk, shaded in blue ($\geq 30\%$ for lifetime risk and $\geq 8\%$ for 10-year risk). Percentages identically equal to zero are denoted as "0" (i.e. no women fall into that category), while percentages less than 0.1 are denoted by "0.0" (i.e. a very small number of women fall into that category).

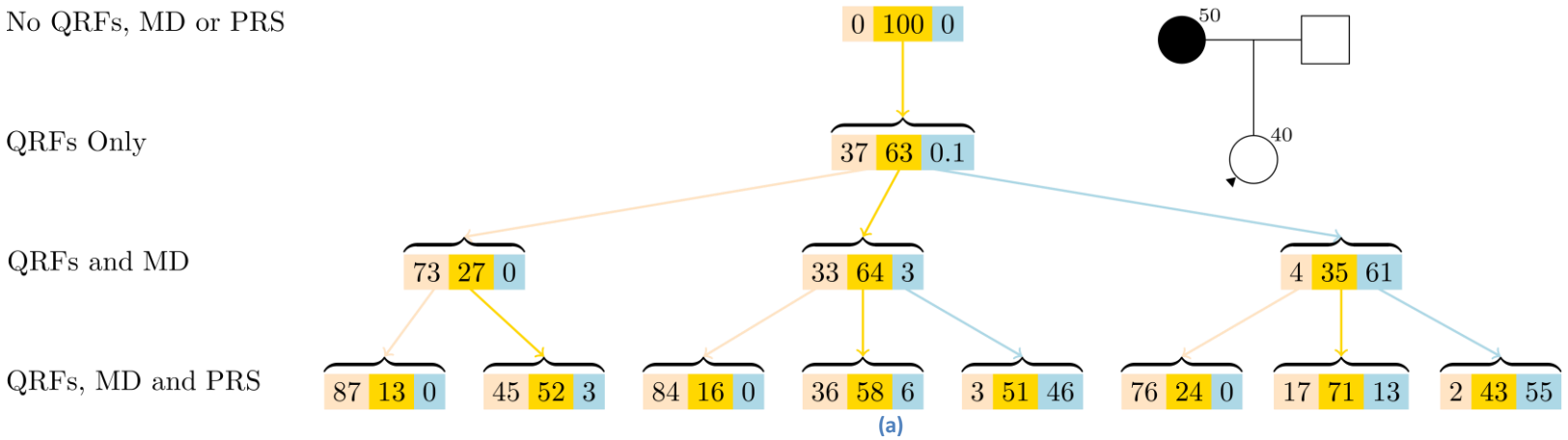


Figure s6.3. Probability Trees for (a) 10-year and (b) lifetime risks, for a woman with a mother affected at age 50, based on BOADICEA. Starting at the top of each tree, the figures show the percent of women reclassified by adding in more information to the breast cancer risk prediction, as indicated by the captions on the left (questionnaire-based risk factors (QRFs), mammographic density (MD) and Polygenic Risk Score (PRS)). Note that in comparison to figure s6.4, here MD is added in before the PRS. Each triplet of numbers is the percentage of women who fall into the familial breast cancer risk categories based on the NICE guidelines[24]: 1) near population risk shaded in pink ($< 17\%$ for lifetime risk and $< 3\%$ for 10-year risk), 2) moderate risk shaded in yellow ($\geq 17\%$ and $< 30\%$ for lifetime risk and $\geq 3\%$ and $< 8\%$ for 10-year risk) and 3) high risk, shaded in blue ($\geq 30\%$ for lifetime risk and $\geq 8\%$ for 10-year risk). Percentages identically equal to zero are denoted as "0" (i.e. no women fall into that category), while percentages less than 0.1 are denoted by "0.0" (i.e. a very small number of women fall into that category).

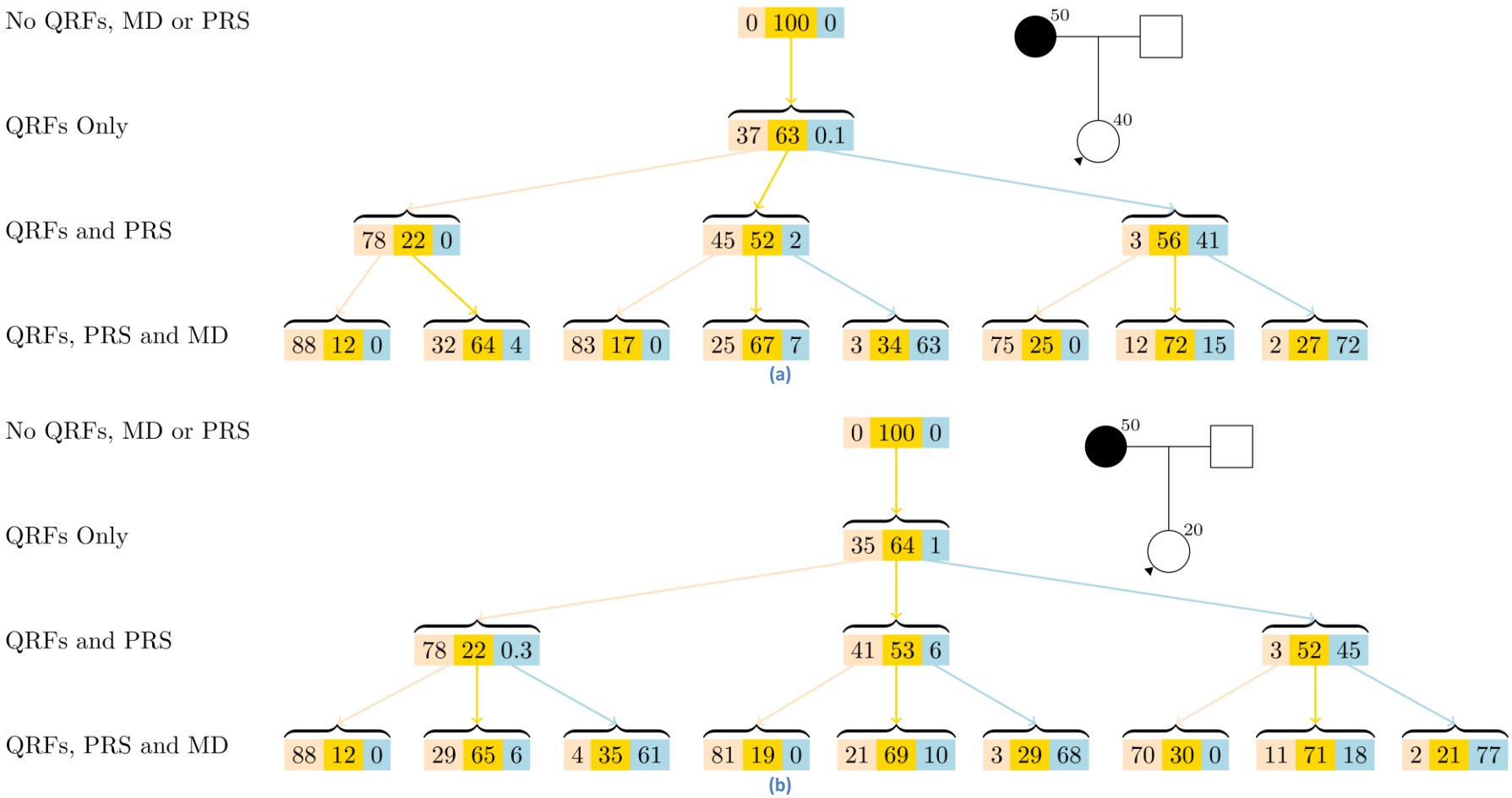


Figure s6.4. Probability Trees for (a) 10-year and (b) lifetime risks, for a woman with a mother affected at age 50, based on BOADICEA. Starting at the top of each tree, the figures show the percent of women reclassified by adding in more information to the breast cancer risk prediction, as indicated by the captions on the left (questionnaire-based risk factors (QRFs), mammographic density (MD) and Polygenic Risk Score (PRS)). Note that in comparison to figure s6.3, here the PRS is added in before MD. Each triplet of numbers is the percentage of women who fall into the familial breast cancer risk categories based on the NICE guidelines[24]: 1) near population risk shaded in pink ($< 17\%$ for lifetime risk and $< 3\%$ for 10-year risk), 2) moderate risk shaded in yellow ($\geq 17\%$ and $< 30\%$ for lifetime risk and $\geq 3\%$ and $< 8\%$ for 10-year risk) and 3) high risk, shaded in blue ($\geq 30\%$ for lifetime risk and $\geq 8\%$ for 10-year risk). Percentages identically equal to zero are denoted as "0" (i.e. no women fall into that category), while percentages less than 0.1 are denoted by "0.0" (i.e. a very small number of women fall into that category).

References

1. Antoniou AC, Cunningham AP, Peto J, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br J Cancer.* 2008;98(8):1457-1466.
2. Macinnis RJ, Antoniou AC, Eeles RA, et al. A risk prediction algorithm based on family history and common genetic variants: application to prostate cancer with potential clinical impact. *Genet Epidemiol.* 2011;35(6):549-556.
3. Lange K. *Mathematical and statistical methods for genetic analysis.* New York: Springer; 1997.
4. Garcia-Closas M, Gunsoy NB, Chatterjee N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. *J Natl Cancer Inst.* 2014;106(11).
5. Pal Choudhury P, Wilcox A, Brook M, et al. Comparative validation of breast cancer risk prediction models and projections for future risk stratification. *bioRxiv.* 2018.
6. 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. *Annals of internal medicine.* 2008;148(5):337-347.
7. Nelson HD, Zakher B, Cantor A, et al. Risk factors for breast cancer for women aged 40 to 49 years: a systematic review and meta-analysis. *Annals of internal medicine.* 2012;156(9):635-648.
8. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev.* 2006;15(6):1159-1169.
9. Swerdlow AJ, Jones ME, Schoemaker MJ, et al. The Breakthrough Generations Study: design of a long-term UK cohort study to investigate breast cancer aetiology. *Br J Cancer.* 2011;105(7):911-917.
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.
11. Office for National Statistics UK. Cohort Fertility Tables in England and Wales. 2011.
12. Reeves GK, Pirie K, Green J, Bull D, Beral V, Million Women Study C. 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.
13. National Centre for Social Research and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2005. 3rd Edition ed: UK Data Service; 2011.
14. National Centre for Social Research and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2006. 4th Edition ed: UK Data Service; 2011.
15. Hunter DJ, Colditz GA, Hankinson SE, et al. Oral contraceptive use and breast cancer: a prospective study of young women. *Cancer Epidemiol Biomarkers Prev.* 2010;19(10):2496-2502.
16. 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.
17. Beral V, Reeves G, Bull D, Green J, Million Women Study C. Breast cancer risk in relation to the interval between menopause and starting hormone therapy. *J Natl Cancer Inst.* 2011;103(4):296-305.
18. Zhang B, Shu XO, Delahanty RJ, et al. Height and Breast Cancer Risk: Evidence From Prospective Studies and Mendelian Randomization. *J Natl Cancer Inst.* 2015;107(11).

19. Green J, Cairns BJ, Casabonne D, et al. Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *The Lancet Oncology.* 2011;12(8):785-794.
20. National Centre for Social Research and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2002. 2nd Edition ed: UK Data Service; 2010.
21. Hamajima N, Hirose K, Tajima K, et al. Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *Br J Cancer.* 2002;87(11):1234-1245.
22. Schmidt MK, Hogervorst F, van Hien R, et al. Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. *J Clin Oncol.* 2016;34(23):2750-2760.
23. Lee AJ, Cunningham AP, Tischkowitz M, et al. Incorporating truncating variants in PALB2, CHEK2, and ATM into the BOADICEA breast cancer risk model. *Genet Med.* 2016.
24. National Institute for Health and Care Excellence. Familial breast cancer: classification and care of people at risk of familial breast cancer and management of breast cancer and related risks in people with a family history of breast cancer (CG164). 2013; http://www.nice.org.uk/CG164
25. Lee AJ, Cunningham AP, Kuchenbaecker KB, Mavaddat N, Easton DF, Antoniou AC. BOADICEA breast cancer risk prediction model: updates to cancer incidences, tumour pathology and web interface. *Br J Cancer.* 2014;110(2):535-545.
26. Waterhouse J, Muir, C.S., Correa and P.,Powell, J., editors. Cancer Incidence in Five Continents, Vol. III (electronic version). IARC Scientific Publication No. 15. Lyon: International Agency for Research on Cancer. (1976) Accessed 2018-06-27; http://ci5.iarc.fr.
27. Waterhouse J, Muir, C.S., Shanmugaratnam, K. and Powell, J., editors Cancer Incidence in Five Continents, Vol. IV (electronic version). IARC Scientific Publication No. 42. Lyon: International Agency for Research on Cancer. (1982) Accessed 2018-06-27; http://ci5.iarc.fr.
28. Muir CS, Waterhouse, J., Mack, T., Powell, J., Whelan and S.L., editors. Cancer Incidence in Five Continents, Vol. V (electronic version). IARC Scientific Publication No. 88. Lyon: International Agency for Research on Cancer. (1987) Accessed: 2018-06-27; http://ci5.iarc.fr.
29. Parkin DM, Muir, C.S., Whelan, S.L., Gao, Y.-T., Ferlay, J. and Powell, J., editors. Cancer Incidence in Five Continents, Vol. VI (electronic version). IARC Scientific Publication No. 120. Lyon: International Agency for Research on Cancer. (1992) Accessed 2018-06-27; http://ci5.iarc.fr.
30. Parkin DM, Whelan, S.L., Ferlay, J., Raymond, L., and Young, J., editors. Cancer Incidence in Five Continents, Vol.VII (electronic version). IARC Scientific Publication No. 143. Lyon: International Agency for Research on Cancer. (1997) Accessed 2018-06-27; http://ci5.iarc.fr.
31. Parkin DM, Whelan, S.L., Ferlay, J., Teppo, L., and Thomas, D.B., editors. Cancer Incidence in Five Continents, Vol. VIII (electronic version). IARC Scientific Publication No. 155. Lyon: International Agency for Research on Cancer. (2002) Accessed 2018-06-27; http://ci5.iarc.fr.
32. Curado MP, Edwards B., Shin H.R., Storm H., Ferlay J., Heanue M. and Boyle P., editors. Cancer Incidence in Five Continents, Vol. IX (electronic version). IARC Scientific Publication No. 160. Lyon: International Agency for Research on Cancer. (2007) Accessed 2018-06-27; http://ci5.iarc.fr.
33. Forman D, Bray F., Brewster D.H., Gombe Mbalawa C., Kohler B., Piñeros M., Steliarova-Foucher E., Swaminathan R. and Ferlay J., editors Cancer Incidence in Five Continents, Vol. X (electronic version). IARC Scientific Publication No. 164. Lyon: International Agency for Research on Cancer. (2014) Accessed 2018-06-27; http://ci5.iarc.fr.
34. Bray F, Colombet M., Mery L., Piñeros M., Znaor A., Zanetti R. and Ferlay J., editors. Cancer Incidence in Five Continents, Vol. XI (electronic version). IARC Scientific Publication No. 166. Lyon: International Agency for Research on Cancer. (2017) Accessed 2018-06-27; http://ci5.iarc.fr.

35. Spanish Network of Cancer Registries (REDECAN). Data supplied by REDECAN. Accessed: 2016-07-15.