Supplementary Material

In this document, we detail the mathematical model used, the form of the likelihood function, the prior distributions for parameters, and the method used to estimate the parameters. Results of our sensitivity analysis are also provided.

Basic natural history model

Let $X_i(t)$ be the disease state of woman i at time t years after the start of the study, with

$$X_i(t) \in \mathcal{X} = \{N, S_{0A}, D_{0A}, S_{0B}, D_{0B}, S_1, D_1, S_2, D_2, S_3, D_3, S_4, D_4\} \tag{1}$$

where

- N indicates no cancer is present;
- S_{0A} and S_{0B} indicate that indolent and aggressive ductal carcinomas in situ (DCIS) are present but not yet diagnosed;
- $S_1, \ldots, S_4$ indicate asymptomatic invasive tumours of size $\leq$ 10mm, 11–20mm, 21–50mm, and $\geq$ 51mm, respectively;
- $D_{0A}, \ldots, D_4$ indicate corresponding symptomatic cancerous states.

We assume exchangeability between women as no risk factors are recorded. For notational simplicity, we will drop the subscript in $X_i(t)$ when not referring to a specific individual. Also, for conciseness, we use θ for the parameter vector containing all unknown parameters.

Let $\pi\left(X(0)|\theta\right)$ be the probability mass for each state in $\mathcal{X}$ at time 0. Since only asymptomatic women meet the entry criteria, all entries in $\pi\left(X(0)|\theta\right)$ corresponding to D states must be 0. Let

$$\pi\left(X(0)|\theta\right) = \begin{pmatrix} \pi_N & \pi_{0A} & 0 & \pi_{0B} & 0 & \pi_1 & 0 & \pi_2 & 0 & \pi_3 & 0 & \pi_4 & 0 \end{pmatrix}^T$$

where $\pi_k \in [0,1]$ is the (unknown) probability of starting in state k and $\pi_N = 1 - \sum_{i=0A}^{4} \pi_i$.

Let the transition rate matrix be $Q(\theta)$ equal to

$$\begin{bmatrix} -\kappa & \kappa(1-\gamma) & 0 & \kappa\gamma & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\delta_0 & \delta_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -(\delta_0+\lambda_0) & \delta_0 & \lambda_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -(\delta_1+\lambda_1) & \delta_1 & \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -(\delta_2+\lambda_2) & \delta_2 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -(\delta_3+\lambda_3) & \delta_3 & \lambda_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\delta_4 & \delta_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Let $T_{01} = 1/12$ be the average time (in years) from the start of the study to the first screen, $T_{02} = 28/12$ from the start to the second screen, and $T_{12} = 27/12$ the inter-screen period, as derived from Fagerberg *et al* (1985).

Using standard properties of Markov processes (Karlin & Taylor, 1975), the probability mass for $X(t)$ can be derived by evaluating

$$p\left(X(t)|\theta\right) = p\left(X(0)|\theta\right) \exp\{tQ(\theta)\}$$

where $\exp\{\cdot\}$ is the matrix exponential function, which we evaluated numerically in `R` using the `expm` package (Goulet, 2012).

In particular, we can derive the probability mass at the first and second screens to be

$$\begin{aligned} u = p\left(X(T_{01})|\theta\right) &= p\left(X(0)|\theta\right) \exp\{T_{01}Q(\theta)\}, \text{ and} \\ v = p\left(X(T_{02})|\theta\right) &= p\left(X(0)|\theta\right) \exp\{T_{02}Q(\theta)\} \end{aligned}$$

respectively. Henceforth, the kth element of u (v) we will denote by u_k (v_k), where the states are ordered as in Equation (1).

In addition, we can derive the distribution of disease status at the second screen, $X(T_{02})$, conditional on disease status at the first screen $X(T_{01})$, by setting

$$W = \exp\{T_{12}Q(\theta)\}.$$

The conditional distribution can then be obtained by reading off the appropriate row of the 13×13 matrix W.

Screening arm: outcome probabilities

In this section, we describe the components of the likelihood function for the screening arm. There are ten outcome scenarios, excluding differences in tumour sizes.

Scenario 1: cancer is detected after randomisation but before screening.

This is possible if the woman had asymptomatic cancer at time 0 that became symptomatic by time T_{01}, or if she had no cancer and a tumour developed and became symptomatic in this time frame (which is very unlikely given the short period of time). In either case, the relevant entry of u is used.

$$\begin{aligned}
\rho_{11} &= \Pr(\text{symptomatic DCIS under Scenario 1}) = u_3 + u_5,\\
\rho_{12} &= \Pr(\text{symptomatic tumour} \leq 10\text{mm under Scenario 1}) = u_7,\\
\rho_{13} &= \Pr(\text{symptomatic tumour 11–20mm under Scenario 1}) = u_9,\\
\rho_{14} &= \Pr(\text{symptomatic tumour 21–50mm under Scenario 1}) = u_{11},\\
\rho_{15} &= \Pr(\text{symptomatic tumour} \geq 51\text{mm under Scenario 1}) = u_{13}.
\end{aligned}$$

(Recall that aggressive and indolent DCIS are not distinguished in the data.)

Scenario 2: no cancer is detected until the time of the first screen, when the woman attends screening and a tumour is found.

Again, this is possible if the woman had asymptomatic cancer at time 0 that was still aymptomatic by time T_{01}, or if she had no cancer and a tumour developed and remained asymptomatic to time T_{01}. Again, the relevant entry of u is used to determine the probability of a cancer of the detected size, along with the sensitivity of the screen, σ, by size, and the probability of attending the first screen (α_1).

$$\begin{aligned}
\rho_{21} &= \Pr(\text{asymptomatic DCIS under Scenario 2}) = \alpha_1(\sigma_1 u_2 + \sigma_2 u_4),\\
\rho_{22} &= \Pr(\text{asymptomatic tumour} \leq 10\text{mm under Scenario 2}) = \alpha_1\sigma_3 u_6,\\
\rho_{23} &= \Pr(\text{asymptomatic tumour 11–20mm under Scenario 2}) = \alpha_1\sigma_4 u_8,\\
\rho_{24} &= \Pr(\text{asymptomatic tumour 21–50mm under Scenario 2}) = \alpha_1\sigma_5 u_{10},\\
\rho_{25} &= \Pr(\text{asymptomatic tumour} \geq 51\text{mm under Scenario 2}) = \alpha_1\sigma_6 u_{12}.
\end{aligned}$$

Scenario 3: cancer is detected after attending first screen, but before the second is scheduled.

This is possible if the woman (i) had asymptomatic cancer by time T_{01} but was undetected, or (ii) had no cancer and a tumour developed, and became symptomatic by time T_{02}. The probability of a cancer of the detected size is determined by the relevant entry of W, along with the probability of attending the first screen (α_1), testing negative ($1 - \sigma$) by size, having cancer of a size no greater than the subsequently detected size (from the relevant entry of u) and tumour growing to the detected size and then being

detected (from the relevant entry in W).

$$\begin{aligned}
\rho_{31} &= \Pr(\text{symptomatic DCIS under Scenario 3}) \\
&= \alpha_1[u_1 W_{1,3} + (1-\sigma_1)u_2 W_{2,3} + u_1 W_{1,5} + (1-\sigma_2)u_4 W_{4,5}], \\
\rho_{32} &= \Pr(\text{symptomatic tumour} \leq 10\text{mm under Scenario 3}) \\
&= \alpha_1 \left[u_1 W_{1,7} + \sum_{k=2}^{3} (1-\sigma_k) u_{2k} W_{2k,7} \right], \\
\rho_{33} &= \Pr(\text{symptomatic tumour 11–20mm under Scenario 3}) \\
&= \alpha_1 \left[u_1 W_{1,9} + \sum_{k=2}^{4} (1-\sigma_k) u_{2k} W_{2k,9} \right], \\
\rho_{34} &= \Pr(\text{symptomatic tumour 21–50mm under Scenario 3}) \\
&= \alpha_1 \left[u_1 W_{1,11} + \sum_{k=2}^{5} (1-\sigma_k) u_{2k} W_{2k,11} \right], \\
\rho_{35} &= \Pr(\text{symptomatic tumour} \geq 51\text{mm under Scenario 3}) \\
&= \alpha_1 \left[u_1 W_{1,13} + \sum_{k=2}^{6} (1-\sigma_k) u_{2k} W_{2k,13} \right].
\end{aligned}$$

Scenario 4: no cancer is detected until the time of the second screen (given that woman attended the first), when the woman attends screening and a tumour is found.

This is possible if (i) the woman had asymptomatic cancer by time T_{01} but was undetected during first screen, and it remained aymptomatic to time T_{02}, or if (ii) she had no cancer at the time of the first screen, a tumour developed and remained asymptomatic to time T_{02}. Again, the probability of this outcome is determined by the probability of attending the two screens (α_1 and α_{21}), the sensitivity of the screens for the one or two tumour sizes ($1-\sigma$), having cancer of a size no greater than the subsequently detected size (relevant entry of u) and tumour growing to the detected size from each possible size at the first screen (from W).

$$\begin{aligned}
\rho_{41} &= \Pr(\text{asymptomatic DCIS under Scenario 4}) \\
&= \alpha_1\alpha_{21}\{\sigma_1[u_1 W_{1,2} + (1-\sigma_1)u_2 W_{2,2}] + \sigma_2[u_1 W_{1,4} + (1-\sigma_2)u_4 W_{4,4}]\}, \\
\rho_{42} &= \Pr(\text{asymptomatic tumour} \leq 10\text{mm under Scenario 4}) \\
&= \alpha_1\alpha_{21}\sigma_3 \left[u_1 W_{1,6} + \sum_{k=2}^{3} (1-\sigma_k) u_{2k} W_{2k,6} \right], \\
\rho_{43} &= \Pr(\text{asymptomatic tumour 11–20mm under Scenario 4}) \\
&= \alpha_1\alpha_{21}\sigma_4 \left[u_1 W_{1,8} + \sum_{k=2}^{4} (1-\sigma_k) u_{2k} W_{2k,8} \right], \\
\rho_{44} &= \Pr(\text{asymptomatic tumour 21–50mm under Scenario 4}) \\
&= \alpha_1\alpha_{21}\sigma_5 \left[u_1 W_{1,10} + \sum_{k=2}^{5} (1-\sigma_k) u_{2k} W_{2k,10} \right], \\
\rho_{45} &= \Pr(\text{asymptomatic tumour} \geq 51\text{mm under Scenario 4}) \\
&= \alpha_1\alpha_{21}\sigma_6 \left[u_1 W_{1,12} + \sum_{k=2}^{6} (1-\sigma_k) u_{2k} W_{2k,12} \right].
\end{aligned}$$

Scenario 5: negative results on both screens.

This is possible if the woman (i) had asymptomatic cancer by time T_{01} that remained asymptomatic to time T_{02} and was undetected during both screens, or (ii) if she had no cancer at the time of the first screen and a tumour developed and remained asymptomatic to time T_{02} and was not detected at the second screen, or (iii) if she were free of the disease for the whole time period. The probability of this scenario is thus obtained by summing over the probabilities of these disjoint events.

$$
\begin{aligned}
\rho_5 &= \Pr\left(\text{Scenario 5}\right) \\
&= \alpha_1\alpha_{21}\Bigg\{v_1 + (1-\sigma_1)u_1W_{1,2} + (1-\sigma_1)^2u_2W_{2,2} \\
&\quad + \sum_{k=2}^{6}(1-\sigma_k)\left[u_1W_{1,2k} + \sum_{j=2}^{k}(1-\sigma_j)u_{2j}W_{2j,2k}\right]\Bigg\}.
\end{aligned}
$$

Scenario 6: no cancer detected in a woman who missed the second screen but attended the first.

This is possible if the woman had asymptomatic cancer by time T_{01} but was undetected during first screen and remained asymptomatic until the end of the study period, or if she were cancer free at the first screen and developed asymptomatic cancer thereafter, or if she were free of the disease for the full duration of the study. Again, the probability of the scenario is derived by summing over these possibilities.

$$
\begin{aligned}
\rho_6 &= \Pr\left(\text{Scenario 6}\right) \\
&= \alpha_1(1-\alpha_{21})\Bigg\{u_1\left(W_{1,1} + \sum_{k=1}^{6}W_{1,2k}\right) + (1-\sigma_1)u_2W_{2,2} \\
&\quad + \sum_{k=2}^{6}\sum_{j=2}^{k}\left[(1-\sigma_j)u_{2j}W_{2j,2k}\right]\Bigg\}.
\end{aligned}
$$

Scenario 7: cancer is detected after missing scheduled first screen, but before the second is scheduled.

This is possible if the woman had asymptomatic cancer by time T_{01} but was undetected as she missed the first screen, or if she had no cancer and a tumour developed and became symptomatic by time T_{02}. The probability of a cancer of the detected size is determined by summing over the all consistent tumour sizes (or lack of a tumour) at the time of the first screen. Note that this is necessary as it is known that that individual did not have symptomatic cancer by that time point, and so we cannot simply inspect the absorbing probabilities in v.

$$
\begin{aligned}
\rho_{71} &= \Pr\left(\text{symptomatic DCIS under Scenario 7}\right) \\
&= (1-\alpha_1)[u_1W_{1,3} + u_2W_{2,3} + u_1W_{1,5} + u_4W_{4,5}] \\
\rho_{72} &= \Pr\left(\text{symptomatic tumour} \leq 10\text{mm under Scenario 7}\right) \\
&= (1-\alpha_1)\left[u_1W_{1,7} + \sum_{k=2}^{3}u_{2k}W_{2k,7}\right] \\
\rho_{73} &= \Pr\left(\text{symptomatic tumour 11–20mm under Scenario 7}\right) \\
&= (1-\alpha_1)\left[u_1W_{1,9} + \sum_{k=2}^{4}u_{2k}W_{2k,9}\right] \\
\rho_{74} &= \Pr\left(\text{symptomatic tumour 21–50mm under Scenario 7}\right) \\
&= (1-\alpha_1)\left[u_1W_{1,11} + \sum_{k=2}^{5}u_{2k}W_{2k,11}\right] \\
\rho_{75} &= \Pr\left(\text{symptomatic tumour} \geq 51\text{mm under Scenario 7}\right) \\
&= (1-\alpha_1)\left[u_1W_{1,13} + \sum_{k=2}^{6}u_{2k}W_{2k,13}\right].
\end{aligned}
$$

Scenario 8: no cancer is detected until the time of the second screen (for a woman who missed the first), when the woman attends screening and a tumour is found.

This is possible if the woman had asymptomatic cancer by time T_{01} but was undetected as she missed the first screen, or if she had no cancer and a tumour developed and remained asymptomatic by time T_{02}. The relevant entry of v is used to determine the probability of an asymptomatic cancer of the detected size, along with the probability of missing the first screen $(1-\alpha_1)$, attending the second screen given that the first was missed (α_{21c}) and sensitivity of the screen (σ) by size.

$$\begin{aligned}
\rho_{81} &= \Pr(\text{asymptomatic DCIS under Scenario 8}) = \alpha_{21c}(1-\alpha_1)[\sigma_1 v_2 + \sigma_2 v_4],\\
\rho_{82} &= \Pr(\text{asymptomatic tumour} \leq 10\text{mm under Scenario 8}) = \alpha_{21c}(1-\alpha_1)\sigma_3 v_6,\\
\rho_{83} &= \Pr(\text{asymptomatic tumour 11–20mm under Scenario 8}) = \alpha_{21c}(1-\alpha_1)\sigma_4 v_8,\\
\rho_{84} &= \Pr(\text{asymptomatic tumour 21–50mm under Scenario 8}) = \alpha_{21c}(1-\alpha_1)\sigma_5 v_{10},\\
\rho_{85} &= \Pr(\text{asymptomatic tumour} \geq 51\text{mm under Scenario 8}) = \alpha_{21c}(1-\alpha_1)\sigma_6 v_{12}.
\end{aligned}$$

Scenario 9: negative results for the second screen, given that the first was missed.

This is possible if the woman had asymptomatic cancer by time T_{02} which was missed on the second screen, or if she were free of the disease. The summation of the relevant entries of v is used to determine the probability of getting negative results for the second screen, along with the probability of missing the first screen $(1-\alpha_1)$, attending the second given that she missed the first (α_{21c}), and testing negative $(1-\sigma)$ if she had asymptomatic cancer.

$$\rho_9 = \Pr(\text{Scenario 9}) = \alpha_{21c}(1-\alpha_1)\left\{v_1 + \sum_{k=1}^{6}[(1-\sigma_k)v_{2k}]\right\}.$$

Scenario 10: missed both screens.

This is possible if the woman had cancer that remained asymptomatic to time T_{02}, or if she were free of the disease.

$$\rho_{10} = \Pr(\text{Scenario 10}) = (1-\alpha_1)(1-\alpha_{21c})\left[v_1 + \sum_{k=1}^{6} v_{2k}\right].$$

Control arm: outcome probabilities

In this section, we describe the components of the likelihood function for the control arm. There are two outcome scenarios, excluding differences in tumour sizes.

Asymptomatic at the end of study.

This is possible if the woman had cancer that remained asymptomatic to time T_{02}, or if she were free of the disease. The summation of the relevant entries of v is used to determine the probability of being asymptomatic under the control arm.

$$\rho_{asym} = \Pr(\text{asymptomatic under control arm}) = v_1 + \sum_{k=1}^{6} v_{2k}.$$

Symptomatic at the end of study.

This is possible if the woman had cancer that became symptomatic by time T_{02}. The relevant entry of v is used to determine the probability of a cancer of the detected size.

$$\begin{aligned}
\rho_{\text{sym1}} &= \Pr(\text{symptomatic DCIS under control arm}) = v_3 + v_5,\\
\rho_{\text{sym2}} &= \Pr(\text{symptomatic tumour} \leq 10\text{mm under control arm}) = v_7,\\
\rho_{\text{sym3}} &= \Pr(\text{symptomatic tumour 11–20mm under control arm}) = v_9,\\
\rho_{\text{sym4}} &= \Pr(\text{symptomatic tumour 21–50mm under control arm}) = v_{11},\\
\rho_{\text{sym5}} &= \Pr(\text{symptomatic tumour} \geq 51\text{mm under control arm}) = v_{13}.
\end{aligned}$$

Outcome probabilities for the 11-state model

The derivation of the outcome probabilities for the model with no indolent cancers can be derived from the above formulation by setting the proportion initially indolent and the proportion of cancers that are indolent to 0.

Derivation of long-run behaviour

The absorption probabilities of D-states and quasi-stationary distributions on the S-states can be obtained in different ways. The first is to insert 'large' values of t in $\Pr(X(t)|\theta) = \Pr(X(0)|\theta)\exp\{tQ(\theta)\}$. The larger t, the closer to the actual long-run behaviour, but the greater the risk of numerical overflow issues in deriving the quasi-stationary distributions. We therefore used a range of values of t to check for consistency.

The second is to reformulate the model as a series of ordinary differential equations determined by Q with an additional 'birth' and 'death' term, μ, that represent influx into the non-cancerous state and efflux from the detected states. The quasi-equilibrium distributions are then obtained by setting each differential equation in the series to 0 and solving, which is straightforward due to their being linear. On doing so, the μ terms cancel. This leads to the same results as the first method when point estimates of the parameters are inserted. In order to obtain the absorption probabilities of the D-states we can use classical theory of Makov chains; see e.g. Karlin & Taylor (1975).

Priors

Informative prior distributions for tumour-dependent screening sensitivity were developed by fitting a binomial distribution to the number of detections from Kerlikowske *et al* (1996) in the relevant size category (Table 3) with an uniform prior for the sensitivity on (0,1). Using Table 2 in Kerlikowske *et al* (1996), we combined the number of cases detected and sample size from the < 50 and ≥ 50 age groups for each tumour size category. We assumed that the sensitivity rates for DCIS and $\geq$ 51mm tumour are similar to that for $\leq$ 10mm and $>$ 20mm tumours respectively. This analysis then yields the following beta posterior distributions that were carried forward as priors:

$$\begin{aligned}
\sigma_j &\sim \text{Be}(77,\, 5), \text{ where } j = 2,\, 3,\\
\sigma_4 &\sim \text{Be}(64,\, 6),\\
\sigma_j &\sim \text{Be}(39,\, 8), \text{ where } j = 5,\, 6.
\end{aligned}$$

Thus, for DCIS we derived a prior distribution using the sensitivity estimates for tumours $\leq$ 10mm, from Kerlikowske *et al* (1996), while for tumours $\geq$ 51mm and tumours 20–50mm, we derived priors using the Kerlikowske *et al* (1996) estimates for tumours $>$ 20mm. This implies that the priors for DCIS and tumours $\leq$ 10mm are the same, though the posteriors need not be, and the priors for tumours 20–50mm and $\geq$ 51mm are the same, though again the posteriors need not be. We did, however, assume $\sigma_1 = \sigma_2$, i.e. that there is no difference in the sensitivity for indolent and aggressive DCIS.

We incorporated external information on the prevalence of indolent DCIS by setting the prior distribution for the probability of getting aggressive breast cancer to be non-informative and incorporating an additional term in the posterior for each screen in which DCIS could be detected, with a parameter characterising the prevalence of indolent DCIS on screening, with an informative prior distribution derived from Leonard & Swain (2004), in which 51 out of 179 detected but untreated DCIS subsequently became invasive.

$$\begin{aligned}
\gamma &\sim \text{U}(0,\, 1),\\
C_1 &\sim \text{Bin}(23,\, \eta),\\
C_2 &\sim \text{Bin}(A_{11},\, \eta) \text{ and}\\
C_3 &\sim \text{Bin}\left((12 - A_{11}),\, \eta\right), \text{ where}\\
\eta &\sim \text{Be}(52,\, 129).
\end{aligned}$$

This allows length-bias to be accounted for.

The prior density for all the transition parameters, κ, ϕ and δ were taken to be Exp(0.01) while the prior density for all other parameters was taken to be $\propto 1$.

To assess the sensitivity to the choice of priors, we performed an alternative analysis in which the point estimates from the Kerlikowske *et al* (1996) and Leonard & Swain (2004) were retained but the sample sizes arbitrarily reduced to half their actual values. The results are presented later in this supplementary material.

Posterior

To prevent overflow issues we work with the natural logarithm of the likelihood and priors. The log-posterior is the summation of the log-likelihood function (incorporating both study and control groups as a summation of the logarithm of two multinomial probability masses, one per arm) and log-priors for screening sensitivity rates, probability of getting aggressive breast cancer and number of DCIS that could be detected for each screen.

Let $\mathcal{N}_{11}$ to $\mathcal{N}_{15}$ be the number of women in the screening arm falling into Scenario 1 with a DCIS or an invasive tumour of size $\leq$ 10mm, 11–20mm, 21–50mm, or $\geq$ 51mm, respectively, $\mathcal{N}_{21}$ to $\mathcal{N}_{25}$, $\mathcal{N}_{31}$ to $\mathcal{N}_{35}$, $\mathcal{N}_{41}$ to $\mathcal{N}_{45}$, $\mathcal{N}_{71}$ to $\mathcal{N}_{75}$, $\mathcal{N}_{81}$ to $\mathcal{N}_{85}$ be the corresponding numbers for Scenarios 2, 3, 4, 7, and 8, and $\mathcal{N}_5$, $\mathcal{N}_6$, $\mathcal{N}_9$, and $\mathcal{N}_{10}$ be the numbers falling into Scenarios 5, 6, 9 and 10, respectively. Similarly, let $\mathcal{M}_0$ be the number of asymptomatic women in the control arm and $\mathcal{M}_1$ to $\mathcal{M}_5$ the number of symptomatic women in these tumour size classes, respectively. Then the log-likelihood function is

$$\begin{aligned}\log \Pr\left(\mathcal{N}, \mathcal{M}|\theta\right) &= c + \sum_{j\in\{1,2,3,4,7,8\}} \sum_{k=1}^{5} \mathcal{N}_{jk} \log \rho_{jk} + \sum_{j\in\{5,6,9,10\}} \mathcal{N}_j \log \rho_j \\ &\quad + \mathcal{M}_0 \log \rho_{a\mathrm{sym}} + \sum_{k=1}^{5} \mathcal{M}_k \log \rho_{\mathrm{sym}\,k}\end{aligned}$$

where c is a constant that factorises in the Metropolis–Hastings step and hence need not be derived. The log-posterior (up to a constant) is the sum of the log-likelihood and the log of the density of the prior.

Proposal distributions

Updates to parameters were proposed within a Markov chain Monte Carlo sampler either to individual parameters or blocks of parameters, followed by the Metropolis–Hastings step. Univariate normal proposal distributions were used for γ, η, κ, α_1, α_{21}, α_{21c}, C_1, C_2 and C_3 and these parameters were accepted or rejected one at a time. Multivariate normal proposal distributions were used for ϕ, δ, σ, A_1 and p_0 and these parameters were accepted or rejected in batches each time. Specifically, using the notation $\psi^\star$ to represent the proposed and ψ^c the current value of a parameter or block of parameters ψ, we proposed parameter values from the following distributions.

Univariate:

$$\begin{aligned}\gamma^\star|\gamma^c &\sim \mathrm{N}(\gamma^c,\, 0.02^2);\\ \eta^\star|\eta^c &\sim \mathrm{N}(\eta^c,\, 0.02^2);\\ \kappa^\star|\kappa^c &\sim \mathrm{N}(\kappa^c,\, 0.00025^2);\\ \alpha_1^\star|\alpha_1^c &\sim \mathrm{N}(\alpha_1^c,\, 0.0015^2);\\ \alpha_{21}^\star|\alpha_{21}^c &\sim \mathrm{N}(\alpha_{21}^c,\, 0.002^2);\\ \alpha_{21c}^\star|\alpha_{21c}^c &\sim \mathrm{N}(\alpha_{21c}^c,\, 0.006^2);\\ C_1^\star|C_1^c &\sim \mathrm{N_D}(C_1^c,\, 1^2);\\ C_2^\star|C_2^c &\sim \mathrm{N_D}(C_2^c,\, 1^2);\\ C_3^\star|C_3^c &\sim \mathrm{N_D}(C_3^c,\, 0.4^2);\end{aligned}$$

where $\mathrm{N_D}$ is a discretised normal.

Multivariate:

$$\phi^\star|\phi^c \sim \mathrm{N}\left(\phi^c, \begin{pmatrix} 10^2 \\ 0.1^2 \\ 0.05^2 \\ 0.05^2 \end{pmatrix} \mathbf{I}_4\right);$$

$$\delta^\star|\delta^c \sim \mathrm{N}\left(\delta^c, \begin{pmatrix} 0.05^2 \\ 0.05^2 \\ 0.05^2 \\ 0.2^2 \\ 1^2 \end{pmatrix} \mathbf{I}_5\right);$$

$$\sigma_j^\star|\sigma_j^c \sim \mathrm{N}\left(\sigma_j^c, \begin{pmatrix} 0.02^2 \\ 0.02^2 \\ 0.02^2 \\ 0.02^2 \\ 0.02^2 \end{pmatrix} \mathbf{I}_5\right); \text{where } j = 2, \ldots, 6 \text{ and } \sigma_1^\star = \sigma_2^\star;$$

$$A_1^\star|A_1^c \sim \mathrm{N_D}\left(A_1^c, \begin{pmatrix} 0.5^2 \\ 2^2 \\ 2^2 \\ 0.7^2 \\ 0.3^2 \end{pmatrix} \mathbf{I}_5\right);$$

$$p_0^\star|p_0^c \sim \mathrm{N}(p_0^c, 0.25 \times \begin{pmatrix} 2446 & -202 & -395 & -660 & -829 & -313 & -47 \\ -202 & 226 & -52 & 18 & 11 & -2 & 2 \\ -395 & -52 & 8375 & -8043 & 94 & 27 & -6 \\ -660 & 18 & -8043 & 8961 & -254 & -28 & 6 \\ -829 & 11 & 94 & -254 & 1016 & -45 & 8 \\ -313 & -2 & 27 & -28 & -45 & 370 & -8 \\ -47 & 2 & -6 & 6 & 8 & -8 & 46 \end{pmatrix} \times 10^{-10}).$$

Run-time parameters were tuned on pilot runs.

Markov chain Monte Carlo

MCMC is a simulation-based approach to sample a complicated distribution of a random variable (Albert, 2007). We could simulate a Markov chain and use its trajectory as the sample from the posterior (Albert, 2007), if a Markov chain with a stationary distribution equal to the posterior could be set up. Metropolis–Hastings algorithm is a common method used to construct a Markov chain that has a specified stationary distribution. To start the running of Markov chain, we first set iteration i to be 1. The initial conditions are arbitrary values chosen to be the initial choices of the parameters to be estimated. We then simulate from the respective proposal distributions, which are functions of two variables—current state of the chain and the candidate value, and evaluate their densities (Albert, 2007). Proposed values, which are drawn from densities of the proposal distributions, are accepted or rejected according to log acceptance probabilities (Albert, 2007). Iteration i is then increased by 1. If the proposed values are rejected, the Markov chain will have a repeat in the sequence. All these lead to correlated samples of the parameters with densities equal to the target densities.

Sensitivity analysis

The parameter estimates from the 11-state model, which does not differentiate indolent and aggressive DCIS are shown in Table 1b in the main paper. There are few differences between corresponding parameters in the two models, with the expectation of a longer sojourn in the DCIS class under the model with no indolent DCIS (0.04 years in 13-state model vs 0.2 years in 11-state model), fewer DCIS shown by the absorption probabilities (9% in 13-state model vs 3% in 11-state model) and more DCIS amongst the steady-state proportions of women with undiagnosed cancer (2% in 13-state model vs 10% in 11-state model). Internal validation indicated the 11-state model (Figure 4 in Supplementary Material) also provided a good description of the data, and there are few differences between their predictions under different mammographic screening frequencies (Figure 5 in Supplementary Material) and breast cancer risks (Figure 6 in Supplementary Material), albeit with fewer predicted ductal carcinomas in situ DCIS in the 11-state model than in the 13-state model.

Sensitivity analysis to prior

We repeated the analysis using weaker informative priors. The results are shown in Table 8 and Figures 8–10. As can be seen from Table 8, most of the parameter and derived parameter estimates are almost unchanged. The most significant change observed is the decrease in average sojourn times for 21–50mm tumour (6.4 years in the main analysis vs 6.1 years in the sensitivity analysis). The model fitted with weaker priors still describes the data well (Figure 8) and the distributions of tumour sizes, regardless of screening intervals (Figure 9) or risk levels (Figure 10), remain similar to that of the original 13-state model.

Annex discussion

We include here various points as there was not enough space to discuss in sufficient length in the main paper.

Tumour size distribution as incidence varies

Although it seems odd intuitively, from a theoretical perspective, the distribution of sizes *must* differ for different baseline incidence rates. The following argument explains why. Imagine two groups, a high and a low risk group, whose incidence is constant from a starting age (say 50) with age within each group. Imagine also that they are intensively monitored over a fixed time horizon, say until the age of 75. Under these assumptions, the distribution of the age at first development of a tumour is exponential (after the starting age of 50, but truncated at 75), with a mean that is smaller for the higher risk group. Since they are monitored only for a fixed time period, there is therefore on average a longer time period for the higher risk women to develop a larger tumour and so the higher risk group should have larger tumours for a given screening frequency.

If, however, both incidence rates are low in absolute terms, the distribution of first tumour development within the monitored age range is approximately uniform, and so the increase in time to develop larger tumours is small. It was not *a priori* clear whether plausible incidence rates, like those considered in this study, would lead to a noticeable difference in tumour sizes, however, the results of our analysis (Figure 3e–3h) suggest that there is almost no such difference.

Limitations of Markov models

As a reviewer points out, an often reported limitation of Markov models is that multiple parameters (in this context growth rates/sojourn times and sensitivities) are difficult to identify jointly in the estimation process, as different combinations of the parameters may have identical, or almost identical, values of the likelihood, i.e. the same ability to explain the data. This may lead to high correlation between the estimates of multiple parameters. This issue reflects a lack of information in a dataset about some characteristics of the model. One solution, and the solution adopted here, is to use Bayesian methods to incorporate information—and the concomitant uncertainty—from other datasets to inform non-identifiable parameters (see Lee *et al* (2011) for an example of this in the context of infectious diseases). In the current context, it is not realistic to expect both progression rates and sensitivities to be determined using data on screening-based diagnoses only, but by incorporating additional data on sensitivities from a different study design, this lack of identifiability can be overcome.

A more fundamental problem with Markov models lies in two assumptions they (typically) make: (i) that rates at which events occur are constant in time, leading to exponentially distributed sojourn times. This assumption provides a good first-order approximation (the mean can be estimated accurately) but not a good second-order approximation (the variance or shape does not agree with how we think the biology should behave: exponential distributions are right-skewed and, we believe, offer a less intuitive interpretation than more common distributions such as normal distributions). We have partially obviated this by distinguishing indolent and aggressive DCIS. (ii) The second, fundamentally related issue is that variability between individuals is not represented within a Markov model. Variability between times *is* but is assumed to be driven entirely by the stochasticity generated by a constant hazard. A consequence is independence between successive events that also serves as a reasonable approximation but does not fully capture biological reality.

Length-biased sampling

One important distinction between our estimates for the proportion of DCIS that are indolent and past estimates is that ours account for length-biased sampling. To those encountering length-biased sampling for the first time, it is counter-intuitive. A popular example in classes on probability illustrates this: if buses arrive randomly at your bus stop and on average 10 minutes apart[1], and you walk to the bus stop, how long do you expect to wait? A common, but wrong, answer is 5 minutes (using the logic that on average you arrive half way between buses). In fact, on average you will wait 10 minutes—because of length-biased sampling. (You're more likely to arrive in a long gap between buses than a short gap.) In the current context, if a tumour is slow growing it is far more likely to be detected while it is in a small pre-invasive state than a faster growing tumour that quickly grows to be larger and invasive is. This means that study designs like that of Leonard & Swain (2004), in which small tumours are selected and proportion that are aggressive determined via follow-up, have to be interpreted with care: they provide estimates not of the fraction of tumours that become aggressive, but of the fraction of *detected* tumours in which the tumour becomes aggressive. Nonetheless, the data they have provided are important, and hence used in our study. To determine the fraction of tumours that ever become aggressive requires knowing something about the speed of progression, i.e. a modelling study such as the current one.

References

- Albert J (2007) *Bayesian Computation with R*. New York: Springer
- Duffy SW, Chen H, Tabar L, Day NE (1995) Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. *Statistics in Medicine* **14**: 1531 – 43
- Duffy SW, Day NE, Tabar L, Chen H, Smith TC (1997) Markov Models of Breast Tumor Progression: Some Age-Specific Results. *Journal of the National Cancer Institute Monographs* **22**
- Fagerberg G, Baldetorp L, Grontoft O, Lundstrom B, Manson JC, Nordenskjold B (1985) Effects of repeated mammographic screening on breast cancer stage distribution: results from a randomized study of 92 934 women in a Swedish county. *Acta Radiol Oncol* **24**(6): 465 – 73
- Goulet V, Dutang C, Maechler M, Firth D, Shapira M, Stadelmann M, Expm-Developers@lists.R-Forge.R-Project.Org (2012) *expm : Matrix exponential*
- Karlin S, Taylor HM (1975) *A First Course in Stochastic Processes*. New York: Academic Press
- Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V (1996) Effect of Age, Breast Density, and Family History on the Sensitivity of First Screening Mammography. *J Am Med Assoc* **276**(1): 33 – 8
- Lee VK, Chen MI, Yap J, Ong J, Lim WY, Lin RTP, Barr I, Ong JBS, Mak TM, Goh LG, Leo YS, Kelly PM, Cook AR (2011) Comparability of Different Methods for Estimating Influenza Infection Rates Over a Single Epidemic Wave. *Am J Epidemiol* **174**(4): 468 – 78
- Leonard GD, Swain SM (2004) Ductal carcinoma in situ, complexities and challenges. *J Natl Cancer Inst* **96**: 906 – 20
- Tabar L, Fagerberg G, Chen H, Duffy SW, Smart CR, Gad A, Smith RA (1995) Efficacy of Breast Cancer Screening by Age. New Results from the Swedish Two-County Trial. *Cancer* **75**(10)
- Wu JC, Hakama M, Anttila A, Yen AM, Malila N, Sarkeala T, Auvinen A, Chiu SY, Chen H (2010) Estimation of natural history parameters of breast cancer based on non-randomized organized screening data: subsidiary analysis of effects of inter-screening interval, sensitivity, and attendance rate on reduction of advanced cancer. *Breast Cancer Res Treat* **122**: 553 – 66

[1] We assume in this description that they arrive as a Poisson process.

Table 2: **List of 36 parameters in breast cancer model.**

Parameter	**Notation**
Breast cancer incidence rate	κ
Probability of getting aggressive breast cancer	γ
Detection rate at tumour size $(i-1)$	δ_i
Progression rate of tumour from size $(i-1)$ to i	ϕ_i
Probability (%) of attending first screening	α_1
Probability (%) of attending second screening, given that woman attended the first	α_{21}
Probability (%) of attending second screening, given that woman did not attend the first	α_{21c}
Sensitivity (%) of mammography for tumour size $(i-1)$	σ_i
Initial probability (%) of being in the thirteen states	p_0
Probability of detected DCIS being aggressive	η
Number of women getting positive second screen results, given that both screens were attended	A_1
Number of women detected with aggressive DCIS in the first screen	C_1
Number of women detected with aggressive DCIS in the second screen, given that the first was attended	C_2
Number of women detected with aggressive DCIS in the second screen, given that the first was missed	C_3

Table 3: **Breast cancer cases detected (x) and subsequently detected (n) by tumour size (Kerlikowske *et al*, 1996).** Data are used for informative prior distribution for screening sensitivity.

Tumour size (mm)	**x**	**n**
DCIS	76	80
≤ 10	76	80
11–20	63	68
21–50	38	45
≥ 51	38	45

Table 4: **Detailed list of parameters and derived parameters estimated by the 13-state Markov model.**

Parameters	**Notation**
Breast cancer incidence rate, per 10,000 woman-years	κ
Indolent breast cancer incidence rate, per 10,000 woman-years	$\kappa(1\text{-}\gamma)$
Aggressive breast cancer incidence rate, per 10,000 woman-years	$\kappa\gamma$
Probability (%) of getting aggressive breast cancer	γ
Initial probability (%)	
(a) No cancer	p_{0_1}
(b) Indolent DCIS	p_{0_2}
(c) Aggressive DCIS	p_{0_4}
(d) $\leq$ 10mm	p_{0_6}
(e) 11–20mm	p_{0_8}
(f) 21–50mm	$p_{0_{10}}$
(g) $\geq$ 51mm	$p_{0_{12}}$
Average sojourn time, in years, for different tumour sizes	
(a) Aggressive DCIS	$\frac{1}{\phi_1}$
(b) $\leq$ 10mm	$\frac{1}{\phi_2}$
(c) 11–20mm	$\frac{1}{\phi_3}$
(d) 21–50mm	$\frac{1}{\phi_4}$
(e) $\geq$ 51mm	$\frac{1}{\delta_5}$
Probability (%) *of detecting breast cancer before progression in tumour size*	
(a) Aggressive DCIS $\rightarrow \leq$ 10mm	$\frac{\delta_1}{\phi_1+\delta_1}$
(b) $\leq$ 10mm $\rightarrow$ 11–20mm	$\frac{\delta_2}{\phi_2+\delta_2}$
(c) 11–20mm $\rightarrow$ 21–50mm	$\frac{\delta_3}{\phi_3+\delta_3}$
(d) 21–50mm $\rightarrow \geq$ 51mm	$\frac{\delta_4}{\phi_4+\delta_4}$
Probability (%) *of attending screenings*	
(a) First screening	α_1
(b) Both screenings	α_{21}
(c) Second screening but not first screening	α_{21c}
Sensitivity (%) *of mammography for different tumour sizes*	
(a) DCIS (indolent or aggressive)	$\sigma_1 = \sigma_2$
(b) $\leq$ 10mm	σ_3
(c) 11–20mm	σ_4
(d) 21–50mm	σ_5
(e) $\geq$ 51mm	σ_6
Number of aggressive DCIS	
(a) Among DCIS detected in first screening	C_1
(b) Among DCIS detected in second screening, given that woman attended the first	C_2
(c) Among DCIS detected in second screening, given that woman did not attend the first	C_3

Table 5: **Parameter estimates and states used in other Markov models.**

Author	States	Estimates
Tabar *et al* (1995)	No detectable disease Preclinical but screening-detectable disease Symptomatic clinical disease	MST=1.7 years (40–49yo) MST=3.3 years (50–59yo) MST=3.8 years (60–69yo) MST=2.6 years (70–74yo)
Duffy *et al* (1995)	No disease Preclinical but detectable disease Clinical disease	MST=2.3 years
Duffy *et al* (1997) — 3-state model	No detectable disease Preclinical screen-detectable disease Symptomatic clinical disease	MST=2.4 years (40–49yo) MST=3.7 years (50–59yo) MST=4.2 years (60–69yo)
Duffy *et al* (1997) — 5-state model	No detectable disease Preclinical node negative disease Preclinical node positive disease Clinical node negative disease Clinical node positive disease	For 50–59yo: λ_1=0.00176 λ_2=0.23 λ_3=0.18 λ_4=0.85
Wu *et al* (2010) — 3-state model	Disease-free Pre-clinical cancer Clinical cancer	MST=2.02 years
Wu *et al* (2010) — 5-state model	Disease-free Pre-clinical localised tumour Pre-clinical non-localised tumour Clinical localised tumour Clinical non-localised tumour	λ_1=0.0025 λ_2=0.3371 λ_3=0.2897 λ_4=1.2230

Table 6: **Informative priors used in the MCMC.**

Prior	Source	Distribution	Point estimate (%)	95% CI (%)	Estimated point estimate (%)	Estimated 95% CI (%)
σ_1,σ_2	Kerlikowske *et al* (1996)	Be (77, 5)	95.0	87.7–98.6	88.3	83.0–92.2
σ_3		Be(77, 5)	95.0	87.7–98.6	90.2	86.5–93.3
σ_4		Be(64, 6)	92.6	83.7–97.6	91.1	87.8–94.0
σ_5		Be(39, 8)	84.4	70.5–93.5	91.9	88.7–94.7
σ_6		Be(39, 8)	84.4	70.5–93.5	93.2	89.9–96.2
η^*	Leonard & Swain (2004)	Be(52, 129)	28.5	22.0–35.7	27.9	21.5–34.7

*If the eight studies were equally weighted, the point estimate (95% CI) for η is 43.0% (33.0–52.5%).

Table 7: **Data used in the study.** Data were extracted from tables and text of Fagerberg *et al* (1985). Ten outcomes of the screening group are numbered in square brackets.

Scenario	**Number of women**
Screening group	
Detected before screen 1 [**1**]	10
DCIS	0
$\leq$ 10mm	2
11–20mm	6
21–50mm	2
$\geq$ 51mm	0
Asymptomatic at screen 1	38 486
Attended screen 1, given that asymptomatic at screen 1	34 168
Positive screen 1 results [**2**]	226
DCIS	23
$\leq$ 10mm	87
11–20mm	79
21–50mm	32
$\geq$ 51mm	5
Negative screen 1 results	33 942
Detected before screen 2, given that negative screen 1 results [**3**]	49
DCIS	4
$\leq$ 10mm	6
11–20mm	20
21–50mm	17
$\geq$ 51mm	2
Asymptomatic at screen 2, given that negative screen 1 results	33 893
Attended screen 2, given that asymptomatic at screen 2 and attended screen 1	29 336
Positive screen 2 results, given that attended screen 1 [**4**]	A1
Negative screen 2 results, given that attended screen 1 [**5**]	B1
Did not attend screen 2, given that asymptomatic at screen 2 and attended screen 1 [**6**]	4 557
Did not attend screen 1, given that asymptomatic at screen 1	4 318
Detected before screen 2, given that did not attend screen 1 [**7**]	22
DCIS	2
$\leq$ 10mm	2
11–20mm	9
21–50mm	8
$\geq$ 51mm	1
Asymptomatic at screen 2, given that did not attend screen 1	4 296
Attended screen 2, given that asymptomatic at screen 2 and did not attend screen 1	745
Positive screen 2 results, given that did not attend screen 1 [**8**]	A2
Negative screen 2 results, given that did not attend screen 1 [**9**]	B2
Did not attend screen 2 [**10**]	3 551
Control group	
Asymptomatic at end	37 659
Symptomatic at end	277

Note: The publication did not distinguish screening results at the second screen by attendance at the first screen: A1+A2=106 where there are 12 DCIS, 39 $\leq$ 10mm, 45 11–20mm, 9 21–50mm, 1 $\geq$ 51mm and B1+B2=29 975.

Table 8: **List of parameter and derived parameter estimates from (a) 13-state model and (b) rerun of model with weaker priors.**

Parameters	**(a) 13-state model**		**(b) With weaker priors**	
	Estimates	**95% CI**	**Estimates**	**95% CI**
Probability (%) of getting aggressive breast cancer	91	85–97	92	85–97
Incidence rate (per 10,000 woman-years)				
(a) Breast cancer	21	17–25	21	17–25
(b) Indolent breast cancer	2	1–3	2	1–3
(c) Aggressive breast cancer	19	16–23	19	16–23
Initial probability (%)				
(a) No cancer	99.23	99.13–99.32	99.23	99.12–99.32
(b) Indolent DCIS	0.06	0.03–0.09	0.06	0.03–0.09
(c) Aggressive DCIS	0.13	0.01–0.32	0.13	0.01–0.32
(d) $\leq$ 10mm	0.21	0.02–0.35	0.21	0.02–0.36
(e) 11–20mm	0.26	0.20–0.33	0.26	0.20–0.33
(f) 21–50mm	0.10	0.07–0.14	0.10	0.07–0.14
(g) $\geq$ 51mm	0.02	0.00–0.03	0.01	0.00–0.03
Average sojourn time, in year, for different tumour sizes				
(a) Aggressive DCIS	0.0	0.0–0.1	0.0	0.0–0.1
(b) $\leq$ 10mm	0.8	0.6–1.1	0.8	0.6–1.1
(c) 11–20mm	2.4	1.6–3.5	2.4	1.6–3.4
(d) 21–50mm	6.4	2.5–15.6	6.1	2.5–14.7
(e) $\geq$ 51mm	0.4	0.1–0.9	0.4	0.1–0.9
Probability (%) of detecting breast cancer before progression in tumour size				
(a) Aggressive DCIS $\rightarrow \leq$ 10mm	0	0–1	0	0–1
(b) $\leq$ 10mm$\rightarrow$11–20mm	12	8–15	12	8–15
(c) 11–20mm $\rightarrow$21–50mm	51	43–60	51	43–59
(d) 21–50mm$\rightarrow \geq$ 51mm	87	79–95	87	78–94
Probability (%) of attending screenings				
(a) First screening	89	88–89	89	88–89
(b) Both screenings	87	86–87	87	86–87
(c) Second screening but not first screening	17	16–19	17	16–19
Sensitivity (%) of mammography for different tumour sizes				
(a) DCIS	88	83–92	87	79–93
(b) $\leq$ 10mm	90	86–93	90	85–94
(c) 11–20mm	91	88–94	92	87–96
(d) 21–50mm	92	89–95	93	89–96
(e) $\geq$ 51mm	93	90–96	95	90–98
Absorption probability of different tumour sizes				
(a) DCIS (indolent or aggressive)	9	4–15	9	4–15
(b) $\leq$ 10mm	10	8–14	10	8–14
(c) 11–20mm	41	35–48	41	35–48
(d) 21–50mm	34	28–41	34	28–41
(e) $\geq$ 51mm	5	2–9	5	2–9
Steady-state proportion of women with undiagnosed cancer				
(a) No cancer	99.4	99.1–99.6	99.4	99.1–99.6
(b) Indolent DCIS	0.2	0.1–0.4	0.2	0.1–0.4
(c) Aggressive DCIS	0.0	0.0–0.0	0.0	0.0–0.0
(d) $\leq$ 10mm	0.1	0.1–0.2	0.1	0.1–0.2
(e) 11–20mm	0.2	0.1–0.3	0.2	0.1–0.3
(f) 21–50mm	0.1	0.0–0.1	0.1	0.0–0.1
(g) $\geq$ 51mm	0.0	0.0–0.0	0.0	0.0–0.0
Steady-state proportion of women with undiagnosed cancer (conditional on some cancer)				
(a) Indolent DCIS	30	12–53	30	12–54
(b) Aggressive DCIS	2	0–4	1	0–4
(c) $\leq$ 10mm	24	15–33	24	15–34
(d) 11–20mm	33	21–47	33	20–46
(e) 21–50mm	10	5–18	10	5–17
(f) $\geq$ 51mm	1	0–2	1	0–2

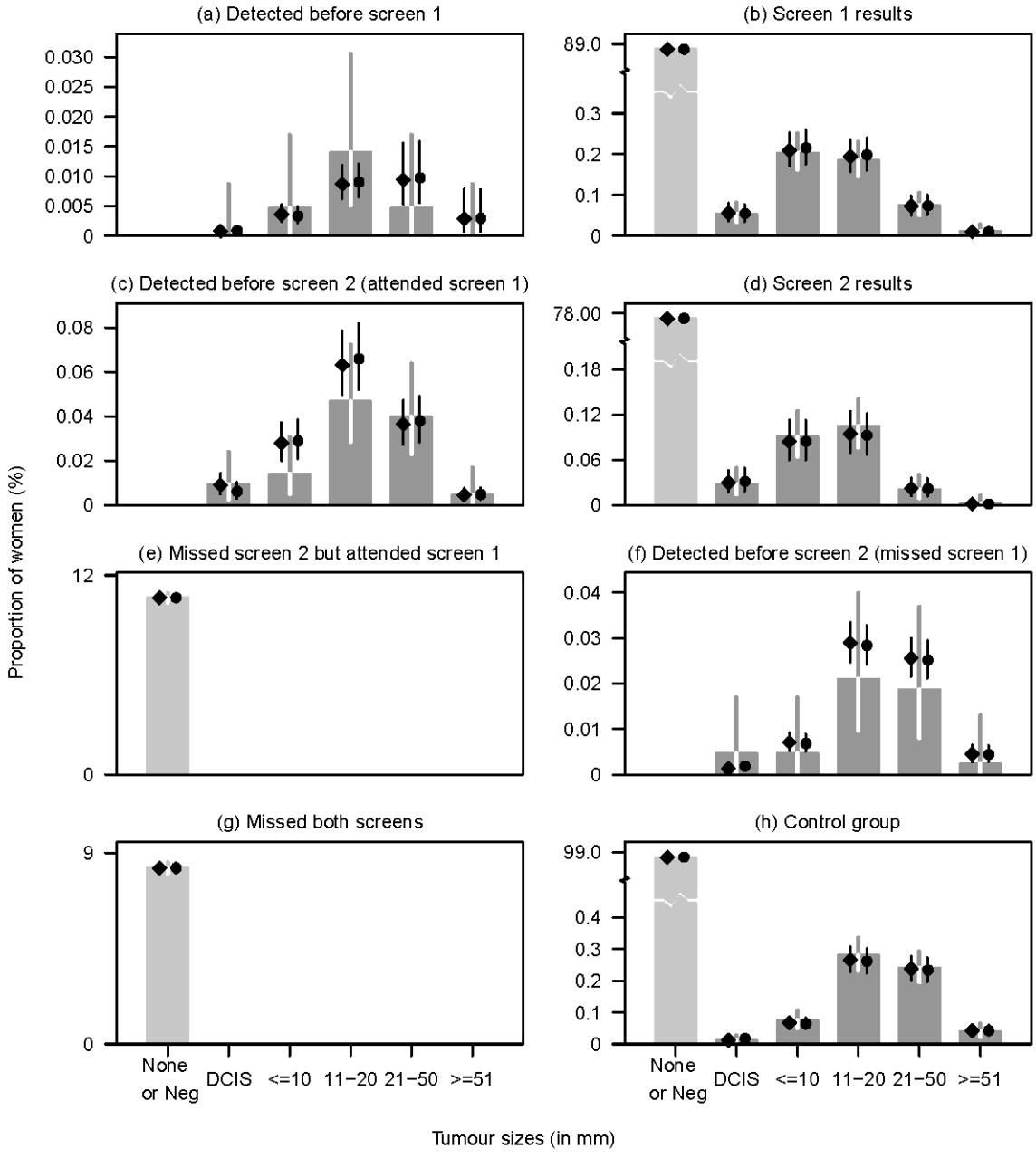


Figure 4: **Data versus predictive distribution of tumour sizes, comparing 11-state and 13-state models.** Bars with lines represent data with their 95% classical confidence intervals based on binomial errors, diamond-shaped points with lines represent modelled proportions and their 95% credible intervals by 11-state model and circular points with lines represent modelled proportions and their 95% credible intervals by 13-state model.

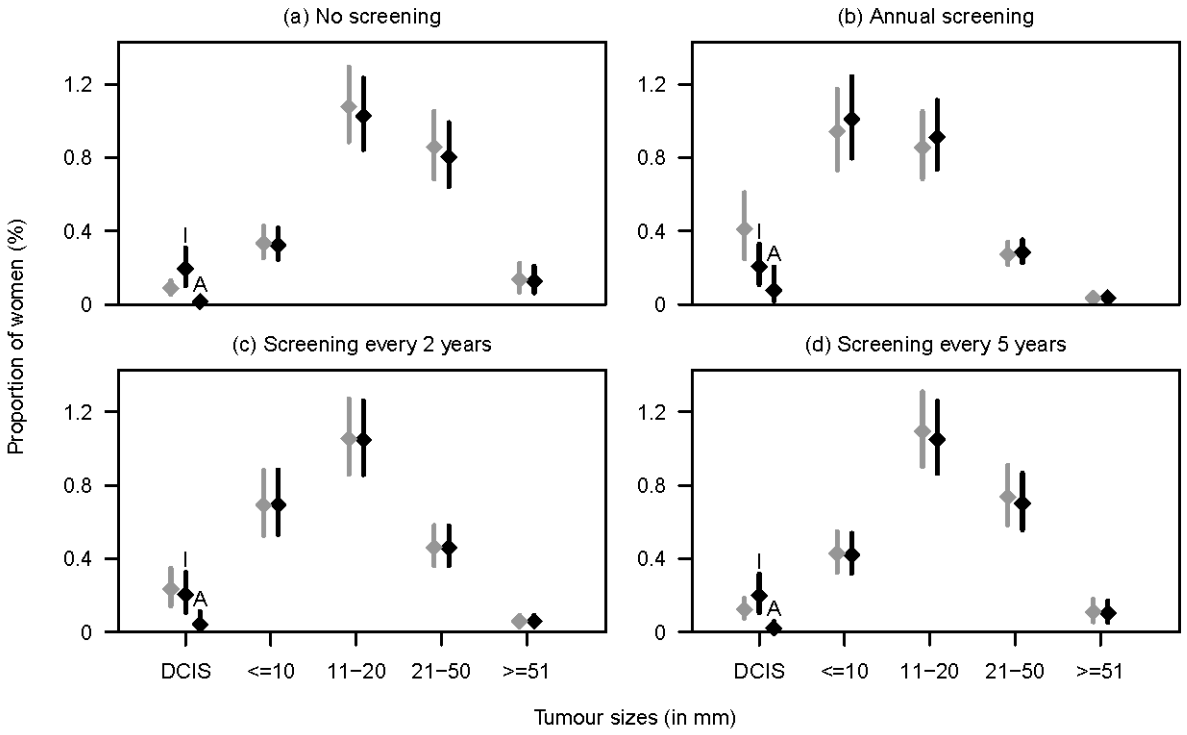


Figure 5: **Tumour size distribution for different mammographic screening frequencies, comparing 11-state and 13-state models.** (a) No screening, (b) annual screening, (c) screening every 2 years, and (d) screening every 5 years. Gray points with lines represent modelled proportions and their 95% credible intervals by 11-state model and black points with lines, to the right, represent modelled proportions and their 95% credible intervals by 13-state model.

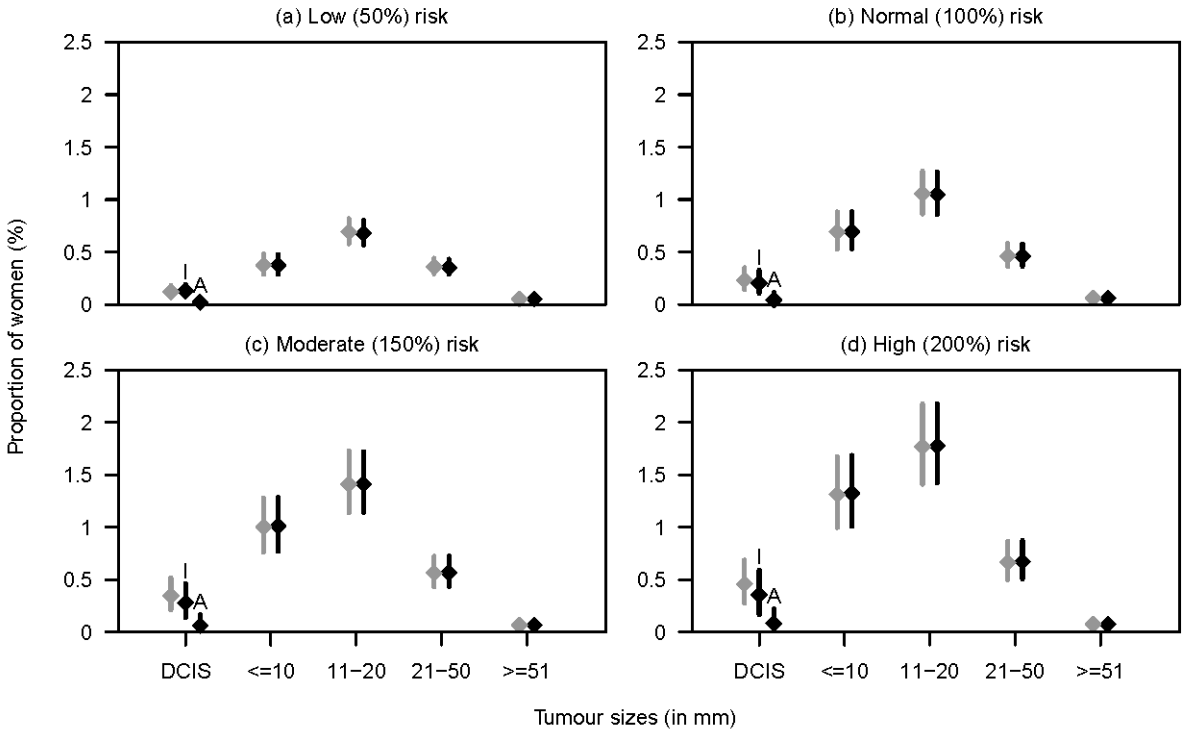
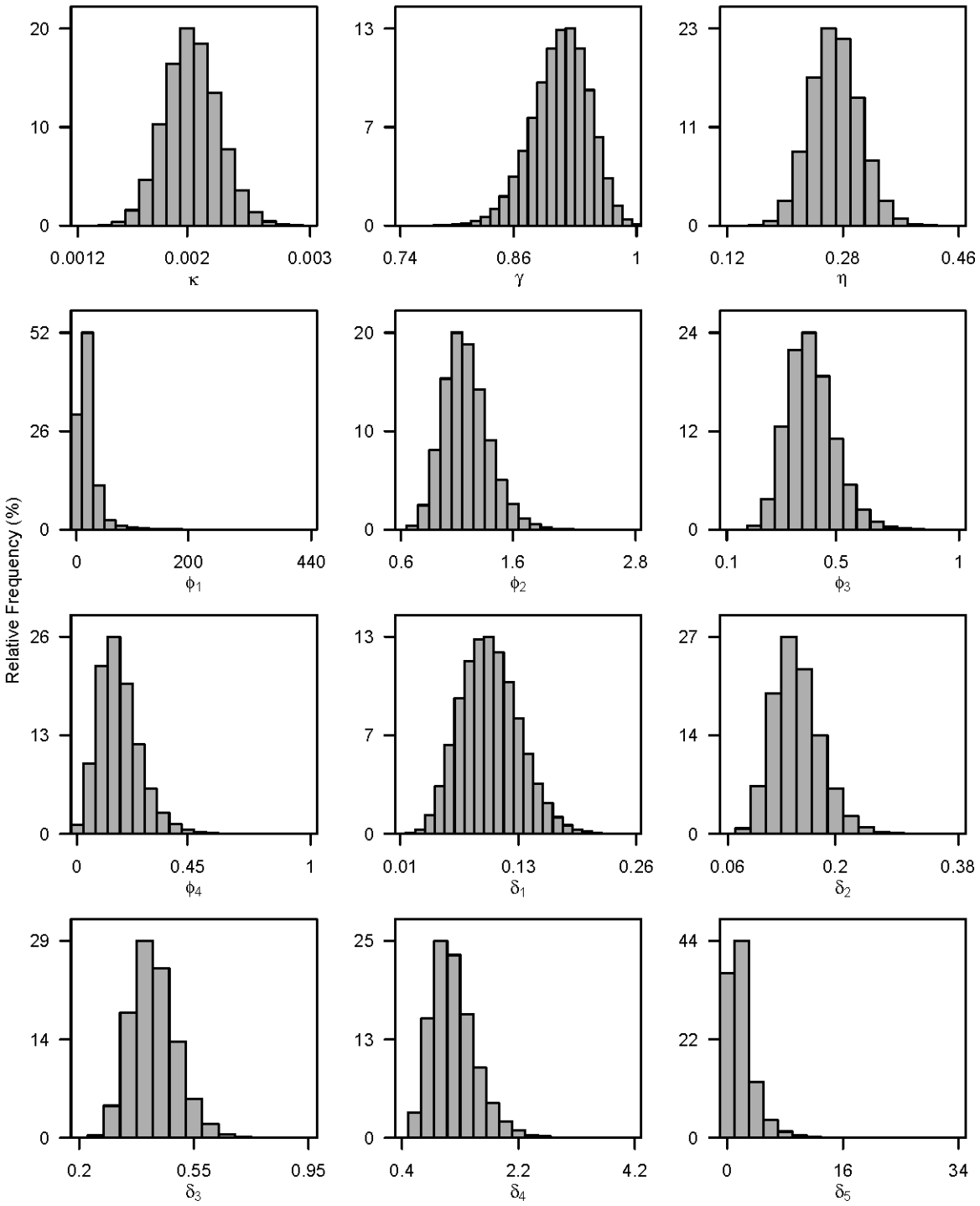
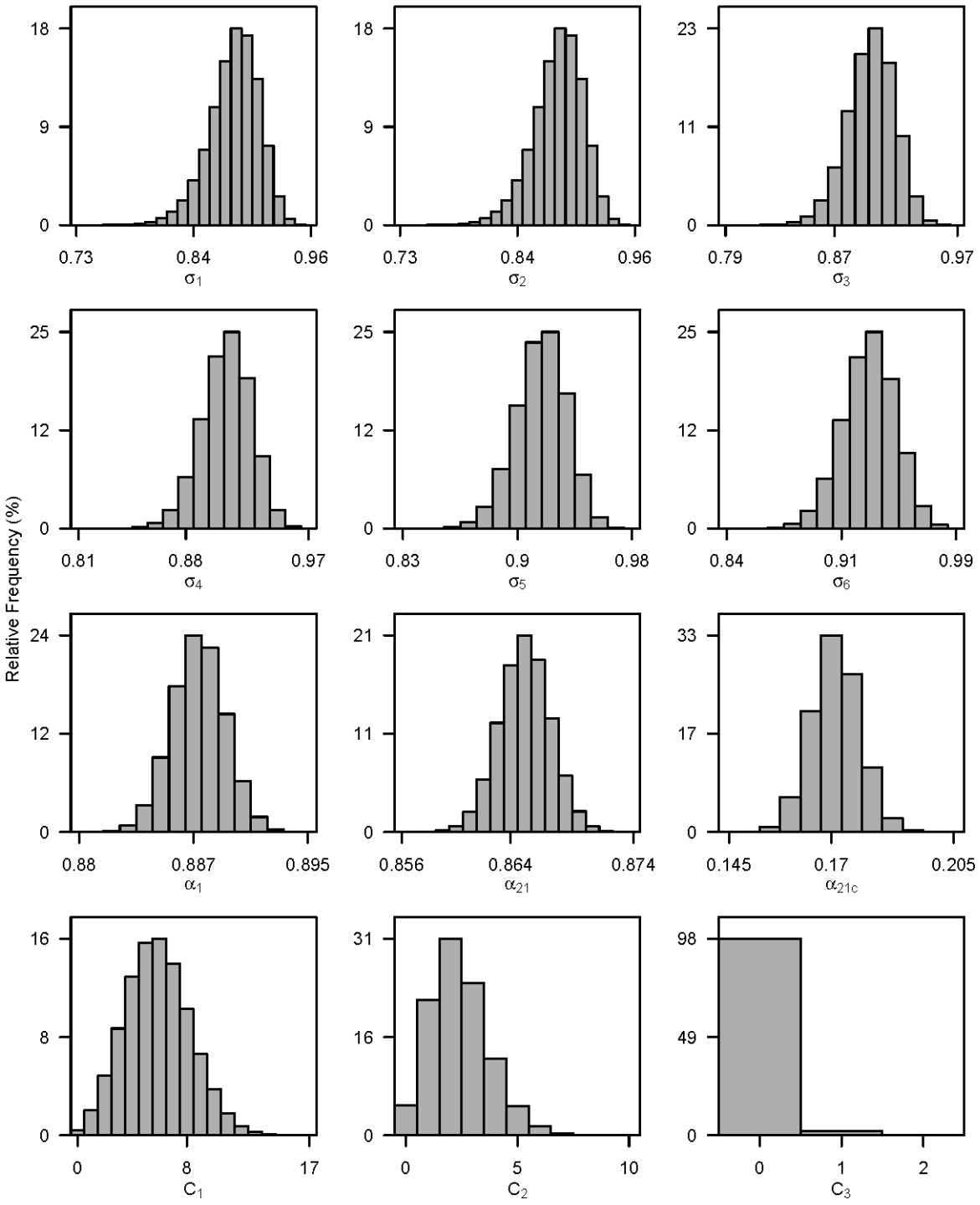


Figure 6: **Tumour size distribution for different rates from no cancer to DCIS, comparing 11-state and 13-state models.** (a) Low risk (50% of baseline), (b) normal risk (100%), (c) moderate risk (150%), and (d) high risk (200%). Gray points with lines represent modelled proportions and their 95% credible intervals by 11-state model and black points with lines, to the right, represent modelled proportions and their 95% credible intervals by 13-state model.


Relative Frequency (%)
κ
γ
η
ϕ_1
ϕ_2
ϕ_3
ϕ_4
δ_1
δ_2
δ_3
δ_4
δ_5


Relative Frequency (%)
σ1
σ2
σ3
σ4
σ5
σ6
α1
α21
α21c
C1
C2
C3

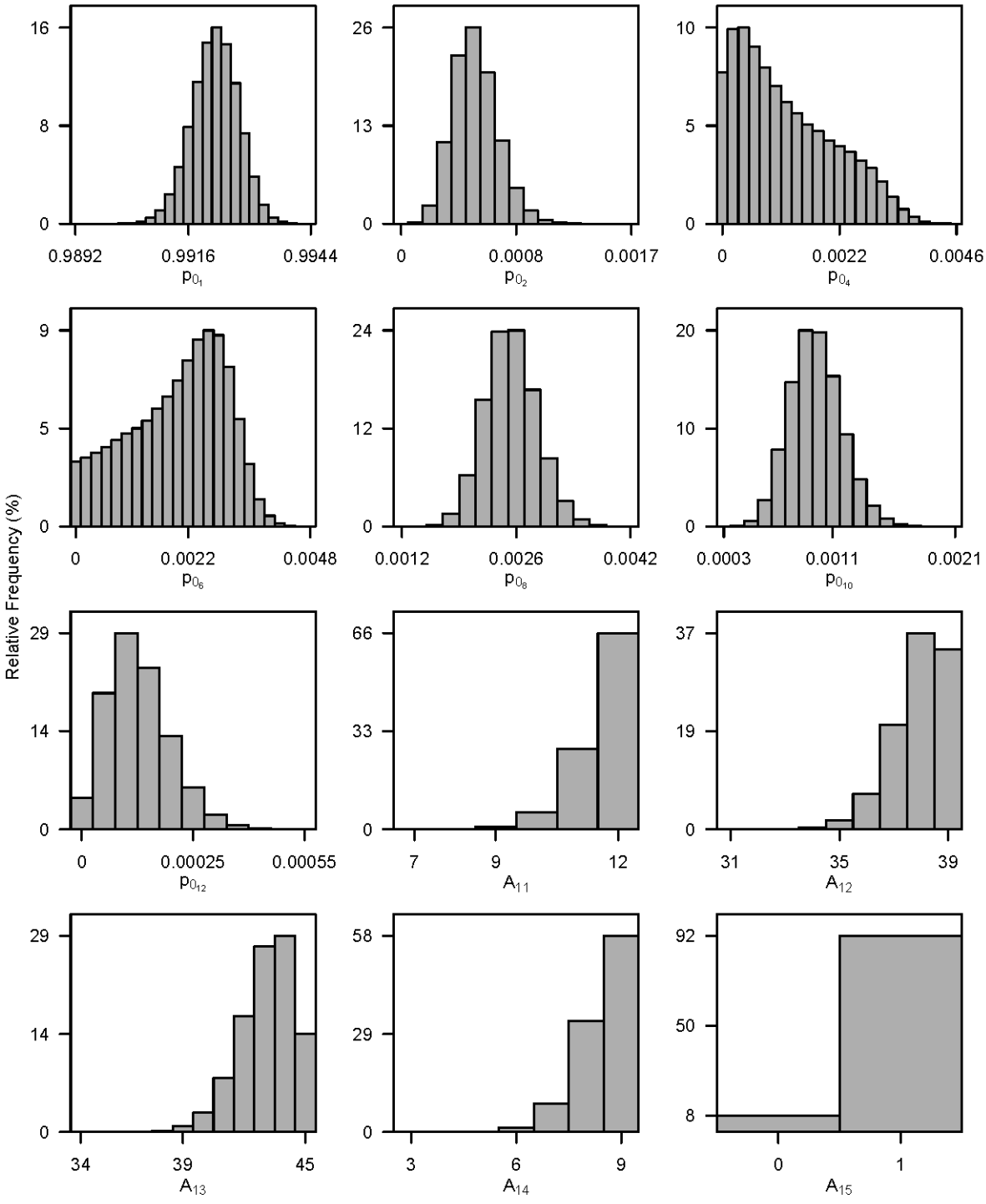


Figure 7: **Histograms of the posterior distributions of estimates.**

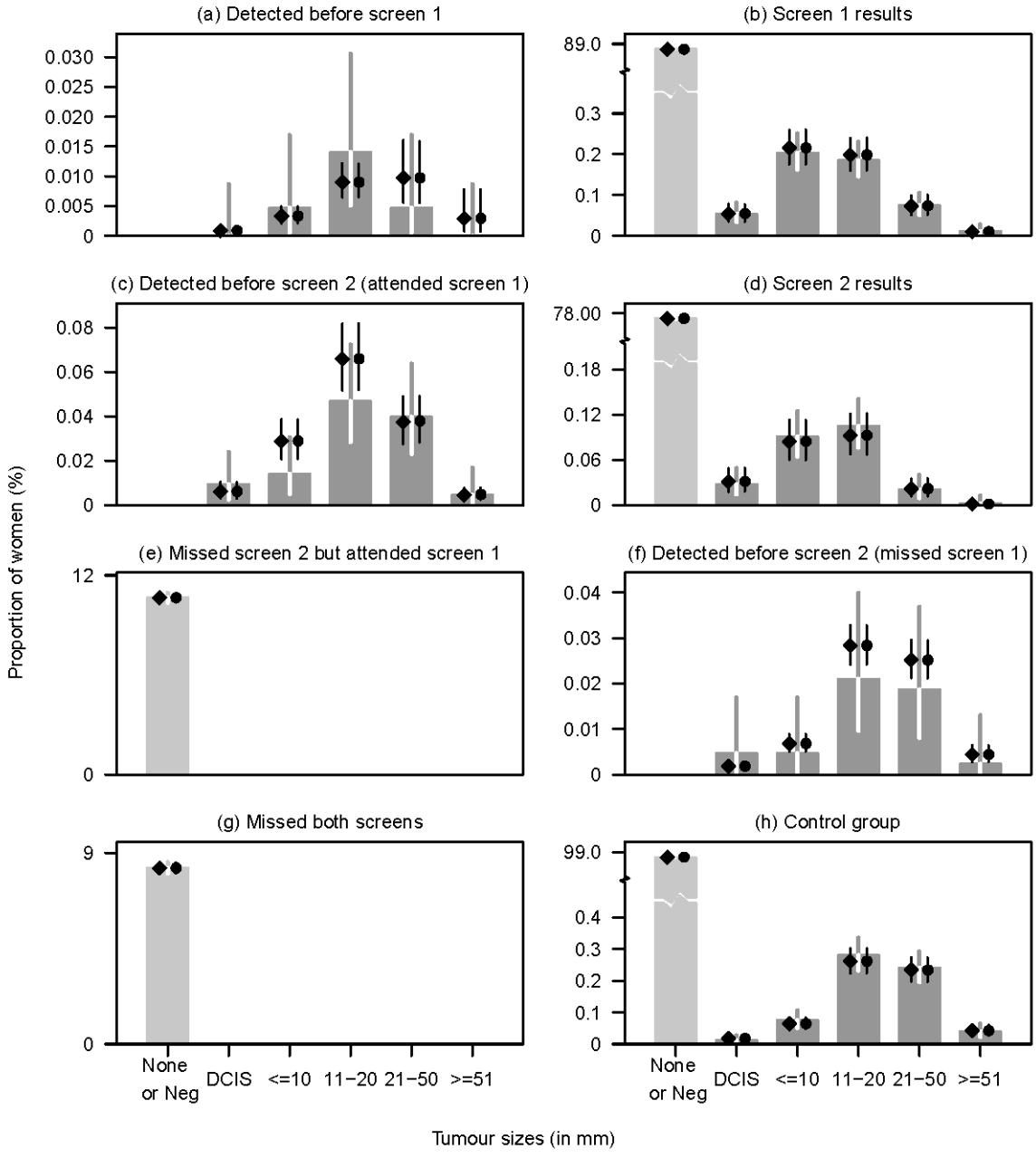


Figure 8: **Data versus predictive distribution of tumour sizes, comparing 13-state model and a rerun of it with weaker priors.** Bars with lines represent data with their 95% classical confidence intervals based on binomial errors, diamond-shaped points with lines represent modelled proportions and their 95% credible intervals by 13-state model with weaker priors and circular points with lines represent modelled proportions and their 95% credible intervals by 13-state model.

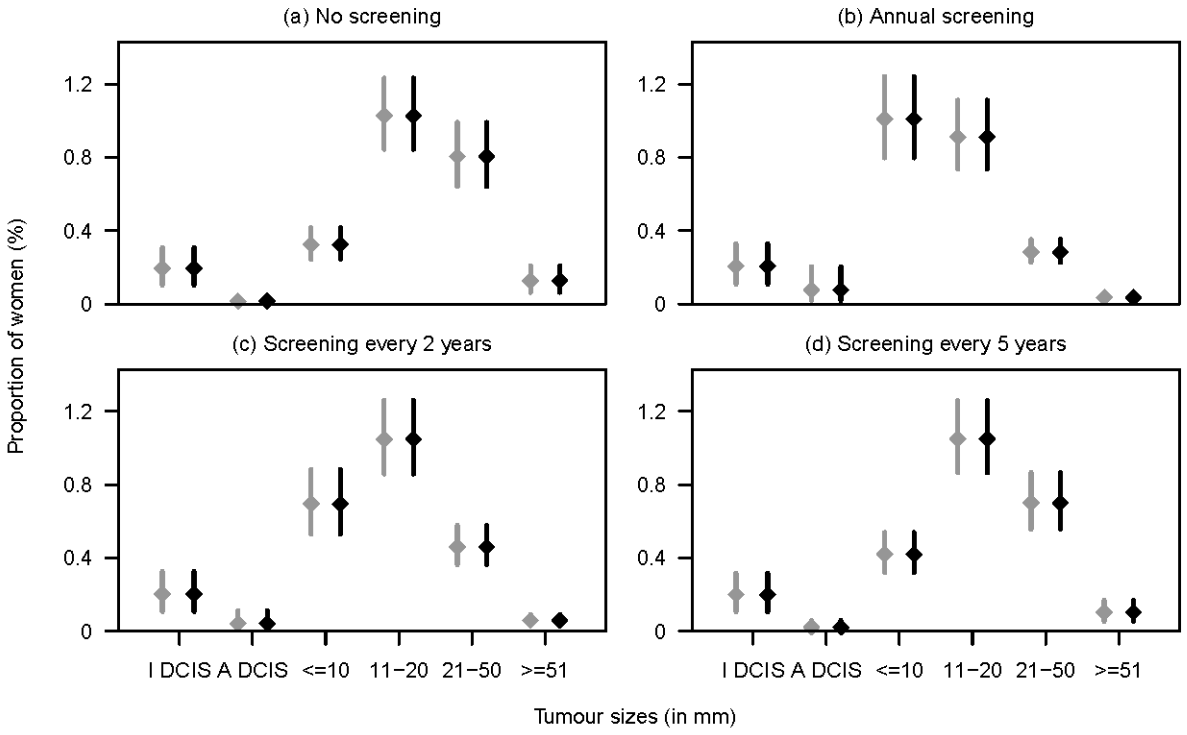


Figure 9: **Tumour size distribution for different mammographic screening frequencies, comparing 13-state model and a rerun of it with weaker priors.** (a) No screening, (b) annual screening, (c) screening every 2 years, and (d) screening every 5 years. Gray points with lines represent modelled proportions and their 95% credible intervals by 13-state model with weaker priors and black points with lines, to the right, represent modelled proportions and their 95% credible intervals by 13-state model.

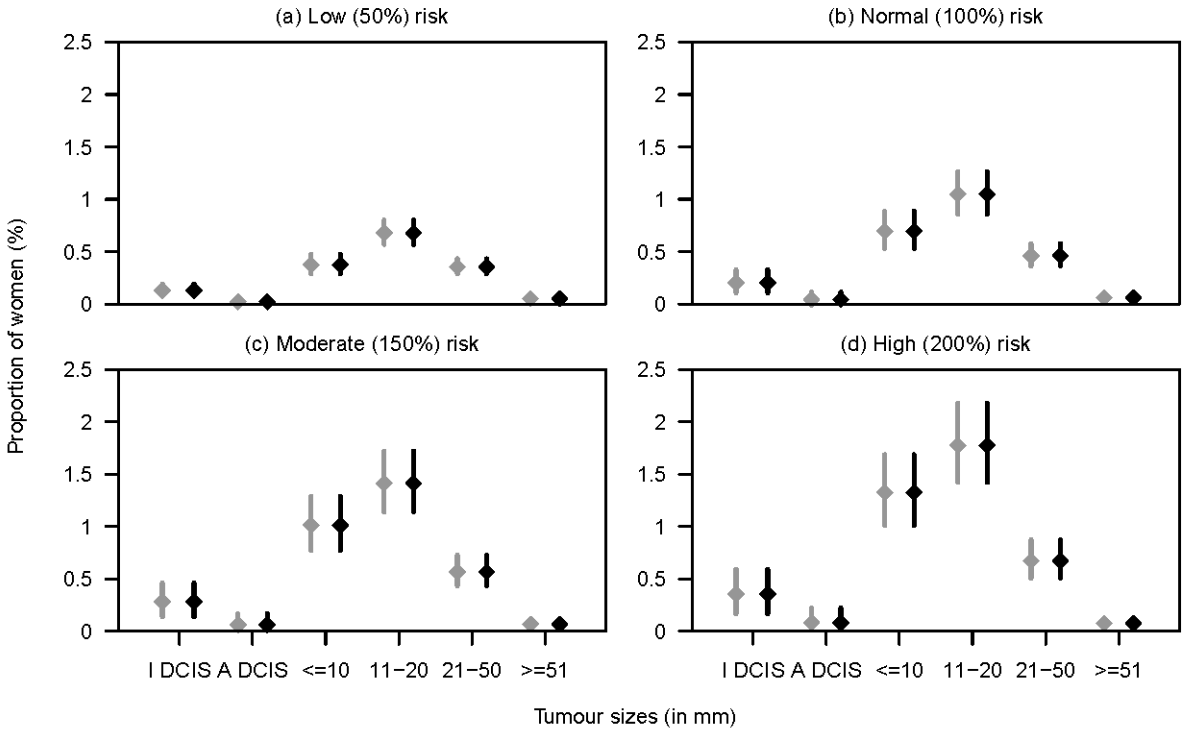


Figure 10: **Tumour size distribution for different rates from no cancer to DCIS, comparing 13-state model and a rerun of it with weaker priors.** (a) Low risk (50% of baseline), (b) normal risk (100%), (c) moderate risk (150%), and (d) high risk (200%). Gray points with lines represent modelled proportions and their 95% credible intervals by 13-state model with weaker priors and black points with lines, to the right, represent modelled proportions and their 95% credible intervals by 13-state model.