WEB APPENDIX

Log-incidence model of breast cancer

We fit our log-incidence model of breast cancer(15) to incident cases of invasive breast cancer that were identified during follow-up of the Nurses' Health Study cohort. The approach to model fitting was to assume that incidence at time *t* (I_t) is proportional to the number of cell divisions (C_t) accumulated throughout life up to age *t*, that is, $I_t = kC_t$. The cumulative number of breast cell divisions is calculated as follows:

$$C_t = C_0 \times \prod_{i=0}^{t-1} (C_{i+1} / C_i) = C_0 \times \prod_{i=0}^{t-1} \lambda_i$$

Thus, $\lambda_i = C_{i+1}/C_i$ represents the rate of increase in the number of breast cell divisions from age *i* to age *i* + 1. Log (λ_i) is assumed to be a linear function of risk factors that are relevant at age *i*. The set of relevant risk factors and their magnitude may vary according to the stage of reproductive life. The details of the representation of C_i have been presented previously.(15) The overall model is given by

$$\begin{aligned}\log I = {} & \gamma_0 + \beta_0 (t^* - t_0) + \beta_1 b + \beta_2 (t_1 - t_0) b_{1,t-1} + \gamma_1 (t - t_m) m_A \\ & + \gamma_2 (t - t_m) m_B + \delta_1 pmh_A + \delta_2 pmh_B + \delta_3 pmh_C + \delta_4 pmh_{cur,t} \\ & + (\delta_4 + \delta_5) pmh_{past,t} + \beta_3 BMI_1 + \beta_3^* BMI_2 + \beta_4 h_1 + \beta_4^* h_2 \\ & + \alpha_1 bbd + \alpha_2 bbd\, t_0 + \alpha_3 bbd (t^* - t_0) + \alpha_4 bbd (t - t_m) m_t \\ & + \phi\, fhx + \beta_5 alc_1 + \beta_6^* alc_2 + \beta_7^{**} alc_3\end{aligned}$$

where *t* = age; t_o = age at menarche; t_m = age at menopause; *t** = minimum (age, age at menopause); m_t = 1 if postmenopausal at age *t*, 0 otherwise; s_t = parity at age *t;* t_i = age at i^{th} birth, *i* = 1,. . , s_t; *b* = birth index = $\sum_{i=1}^{s_t} (t^* - t_i) b_{it}$; b_{it} = 1 if parity ≥ *i* at age *t*, 0 otherwise; m_A= 1 if natural menopause, 0 otherwise; m_B = 1 if bilateral oophorectomy, 0 otherwise; *bbd* = 1 if benign breast disease = yes, 0 otherwise; *fhx* = 1 if family history of

breast cancer in mother or sister = yes, 0 otherwise; pmh_A = number of years on oral estrogen; pmh_B = number of years on oral estrogen and progestin; pmh_C = number of years on other types of postmenopausal hormones; $pmh_{cur,t}$ = 1 if current user of postmenopausal hormones at age *t*, 0 otherwise; $pmh_{past,t}$ = 1 if past user of postmenopausal hormones at age *t*, 0 otherwise; BMI_j = BMI at age *j* (kg/m^2); alc_j = alcohol use (grams) at age *j; h* = height (inches).

β_o represents the rate of increase in incidence before menopause among nulliparous women with no benign breast disease and no family history. β_1 and β_2 represent modifications to the rate of increase in incidence for parous women according to the number and precise spacing of births. γ_1 and γ_2 represent rates of increase in incidence after menopause according to type of menopause among women without benign breast disease not currently using postmenopausal hormones. δ_1, δ_2, and δ_3 represent modifications to the rate of increase in incidence after menopause among women currently using postmenopausal hormones according to the duration of the specific types of postmenopausal hormones used. δ_4 and δ_5 represent the immediate effect of starting and stopping postmenopausal hormone use on rates of increase in incidence after menopause. φ represents the effect of family history of breast cancer on the number of breast cell divisions at birth (i.e., C_o).

The terms for BMI, height, and alcohol use in relation to menopause and postmenopausal use of hormones are summarized below:

$$BMI_1 = \sum_{j=t_0}^{t^*-1} (BMI_j - 21.8) + \sum_{j=t_m}^{t-1} (BMI_j - 24.4)\, pmh_{cur,j} m_j$$

$$BMI_2 = \sum_{j=t_m}^{t-1} (BMI_j - 24.4)(1 - pmh_{cur,j}) m_j$$

$$h_1 = (h - 64.5)(t^* - t_0) + (h - 64.4) \sum_{j=t_m}^{t-1} pmh_{cur,j} m_j$$

$$h_2 = (h - 64.4) \sum_{j=t_m}^{t-1} (1 - pmh_{cur,j} m_j)$$

$$alc_1 = \sum_{j=18}^{t^*-1} alc_j$$

$$alc_2 = \sum_{j=t_m}^{t-1} alc_j\, pmh_{cur,j} m_j$$

$$alc_3 = \sum_{j=t_m}^{t-1} alc_j (1 - pmh_{cur,j}) m_j$$

β_3 (β_4) represents the effect of BMI (height) on breast cancer incidence either before menopause or after menopause while currently using postmenopausal hormones. β_3^* represents the effect of BMI (height) on breast cancer incidence after menopause while not using postmenopausal hormones. β_5, β_5^*, and β_5^{**} represent the effects of alcohol before menopause, after menopause while currently using postmenopausal hormones, and after menopause while not using postmenopausal hormones, respectively. The rationale for the separate terms is the finding in exploratory analyses driven by previous literature(35) that 1) the effects of BMI and possibly height and alcohol use on breast cancer incidence are different before and after menopause and 2) the effect of BMI on breast cancer incidence after menopause differs according to whether a woman is or is not currently using postmenopausal hormones.(36) α_1, α_2, α_3, and α_4 represent modifications, among women with benign breast disease, to 1) the number of breast cell divisions at birth, 2) the rates of increase in the number of cell divisions after birth but

before menarche, 3) the rates of increase in the number of cell divisions after menarche but before menopause, and 4) the rates of increase in the number of cell divisions after menopause. The rationale for the extra terms involving benign breast disease ($\alpha_1, \ldots, \alpha_4$) is that the relative risk for benign breast disease varies according to age, is strongest among younger women, and diminishes over time.

The general rationale for a log-incidence model of a specific cancer is that the number of precancerous cells increases multiplicatively with time, but that the risk factor profile from birth through current age differentially affects the rate of increase in incidence. Specifically, in the breast cancer incidence model described above, the number of precancerous cells is assumed to increase annually at the rate of $\exp(\beta_0)$ before menopause for nulliparous women, at the rate of $\exp(\beta_0 + \beta_1 s)$ before menopause for parous women with parity = s, and so forth. Finally, the number of precancerous cells increases immediately after the first birth by $\exp[\beta_2(t_1 - t_0)]$. The incidence rate of breast cancer is assumed to be approximately proportional to the number of precancerous cells.

The log-incidence model was fit using iteratively reweighted least squares, with PROC NLIN of SAS.(18) The parameters of the model are readily interpretable in a relative risk (RR) context. For example, $\exp(-\beta_0)$ = RR for a 1-year increase in age at menarche among nulliparous women, $\exp[-(\beta_0 + \beta 2)]$ = RR for a 1-year increase in age at menarche among parous women, and so forth.