Supplemental Materials: Detailed descriptions of the FHCRC and UMICH prostate cancer natural history models and estimation methods

1. FHCRC model

The FHCRC model (Gulati et al., 2010, 2013) is designed to simulate a population of men with prostate-specific antigen (PSA) levels and prostate cancer development, progression, and detection patterns that are representative of the general US male population. The model explicitly represents individual PSA growth trajectories, with faster PSA growth associated with shorter times to cancer progression and detection. Details of the PSA growth and the natural/clinical history model components are described below.

1.1 PSA growth

1.1.1 Description

Log PSA for the i th individual at age t (offset from age 35) is:

$$\log[y_i(t)] = \beta_{0i} + \beta_{1i}t + \beta_{2i}(t - t_o)I(t > t_o) + \varepsilon_i(t)$$

where t_o represents age at onset of a preclinical biopsy-detectable tumor (described below), $I(\cdot)$ is the usual indicator function, and the parameters are drawn from normal (N) and truncated normal (N^+) distributions:

$$\beta_{0i} \sim N(\mu_0, \sigma_0^2)$$
$$\beta_{1i} \sim N^+(\mu_1, \sigma_1^2)$$
$$\beta_{2i} \sim \begin{cases} N^+(\mu_{2,2-7}, \sigma_{2,2-7}^2) & \text{Gleason score } 2-7 \\ N^+(\mu_{2,8-10}, \sigma_{2,8-10}^2) & \text{Gleason score } 8-10 \end{cases}$$
$$\varepsilon_i(t) \sim N(0, \tau^2)$$

and the mean and variance meta-parameters have normal and inverse Gamma prior distributions:

$$\mu_k \sim N(0,1) \qquad k = 0,1,(2,2-7),(2,8-10)$$
$$\sigma_k^2 \sim \text{Inverse Gamma}(1,1) \qquad k = 0,1,(2,2-7),(2,8-10)$$
$$\tau^2 \sim \text{Inverse Gamma}(1,1)$$

1.1.2 Estimation

Components of the log-linear PSA growth changepoint model were estimated using fairly non-informative priors and separate linear random effects models for healthy men, for men with Gleason score 2–7 cancer, and for men with Gleason score 8–10 cancer using the most recent 2–4 PSA results before diagnosis from the placebo arm of the Prostate Cancer Prevention Trial. Components were fit using WinBUGS (Medical Research Council Biostatistics Unit, Cambridge, UK, http://www.mrc-bsu.cam.ac.uk/bugs). Following diagnostic evaluation of results using the R package *coda*, parameter estimates are based on means of 5,000 posterior observations (sampling every 10th) for healthy men and 30,000 posterior observations (sampling every 50th) for men diagnosed with cancer after discarding 10,000 burn-in iterations.

We previously combined estimates from an earlier specification not conditioned on Gleason grade with an earlier version of the natural/clinical history model. Based on comparisons between observed and predicted test-positive and cancer detection rates in a simulation of the Prostate, Lung, Colorectal, and

Ovarian cancer screening trial, we modified estimates of the mean log PSA intercept at age 35 and slope before onset (Gulati et al., 2010). However, combining estimates of Gleason grade-dependent post-onset PSA growth with the current version of the natural/clinical history model (described below), only the estimate of the mean log PSA intercept at age 35 was modified from –1.2720 to –1.6094. Final estimates are shown in Table 1.1.

Table 1.1. Results of PSA growth model estimation

Parameter	Estimate
Mean log PSA intercept at age 35 (μ_0)	−1.6094
Mean log PSA slope before onset (μ_1)	0.0446
Mean log PSA slope after onset of Gleason score 2–7 cancer ($\mu_{2,2-7}$)	0.0397
Mean log PSA slope after onset of Gleason score 8–10 cancer ($\mu_{2,8-10}$)	0.1678
SD of log PSA intercept at age 35 (σ_0)	0.2384
SD of log PSA slope before onset (σ_1)	0.0430
SD of log PSA slope after onset of Gleason score 2–7 cancer ($\sigma_{2,2-7}$)	0.0913
SD of log PSA slope after onset of Gleason score 8–10 cancer ($\sigma_{2,8-10}$)	0.3968
SD of log PSA noise (τ)	0.2737

1.2 Natural/clinical history

1.2.1 Description

The hazard of onset of a preclinical biopsy-detectable tumor increases with age according to a Weibull distribution:

$$\lambda_o(t) = \frac{\phi}{\psi}\left(\frac{t}{\psi}\right)^{\phi-1}$$

The probability of Gleason score 8–10 cancer at onset increases quadratically with age at onset:

$$\Pr(\text{Gleason score } 8-10) = \theta_0 + \theta_1 t_o + \theta_2 t_o^2$$

Hazards of metastasis (i.e., progression from stage M0 to M1) and of non-screen diagnosis (i.e., progression from a preclinical to a clinical or symptomatic state) increase with individual PSA:

$$\lambda_m(t) = \gamma \overline{y}_i(t)$$

$$\lambda_c(t) = \begin{cases} \chi_{L,2-7}\overline{y}_i(t) & \text{Localized Gleason score } 2-7 \\ \chi_{L,8-10}\overline{y}_i(t) & \text{Localized Gleason score } 8-10 \\ \chi_{M,2-7}\overline{y}_i(t) & \text{Metastatic Gleason score } 2-7 \\ \chi_{M,8-10}\overline{y}_i(t) & \text{Metastatic Gleason score } 8-10 \end{cases}$$

where $\overline{y}_i(t)$ represents noise-free PSA at time t for the ith individual. This formulation of the natural/clinical history model reflects several changes since the version described in Gulati et al. (2010). Specifically, the hazard of onset now follows a 2-parameter instead of a 1-parameter Weibull distribution, the probability of high Gleason score increases quadratically instead of proportionally with age at onset, and both (a) post-onset PSA growth and (b) stage-specific hazards of clinical diagnosis depend on Gleason score.

1.2.2 Estimation

The model simulates a population of men aged 50–84 in each year in the period 1975–2000 to match counts of men in the core 9 catchment areas of the Surveillance, Epidemiology, and End Results (SEER) program. Simulated men are assigned ages at other-cause death from US life tables (National Center for Health Statistics, various years), PSA growth rates from the PSA growth model, and ages at onset, at transition from localized to metastatic stage, and at clinical diagnosis from probability integral transforms of the natural/clinical history hazard functions. Men are also assigned PSA screening schedules using screening dissemination patterns retrospectively reconstructed by Mariotto et al. (2007) based on National Health Interview Survey responses and SEER-Medicare claims. At each screen, noisy PSA is observed, men with PSA > 4 ng/mL receive biopsy based on age- and PSA-dependent compliance frequencies from the PLCO trial (Pinsky et al., 2005), and a biopsy can detect a latent cancer with sensitivity that increases through the 1990s as biopsy schemes evolved to include an increasing number of cores (see Gulati et al., 2010, and cited references).

Total prostate cancer incidence due to screening and clinical presentation in the simulated population is compared with observed data from SEER by age, year, stage, and grade in a Poisson likelihood:

$$L(\Gamma) = \prod_{s \in S} \frac{e^{-f(\Gamma)} f(\Gamma)^k}{k!}$$

where $\Gamma = (\phi, \psi, \theta_0, \theta_1, \theta_2, \gamma, \chi_{L,2-7}, \chi_{L,8-10}, \chi_{M,2-7}, \chi_{M,8-10})$ is the vector of natural/clinical history parameters, $f(\Gamma)$ is simulated incidence counts, k is observed incidence counts, and the product is across strata S defined by combinations of single-year ages $50, \ldots, 84$ (inclusive); years $1975, \ldots, 2000$ (inclusive); SEER stages local-regional and distant; and SEER grades I-II and III-IV, which we converted to Gleason scores 2–7 and 8–10. Missing stage and/or grade variables were imputed assuming they were missing at random. Results are shown in Table 1.2.

Table 1.2. Results of natural/clinical history model estimation

Parameter	Estimate
Hazard of onset—Weibull scale parameter (ϕ)	0.0012
Hazard of onset—Weibull shape parameter (ψ)	1.6499
Probability of Gleason score 8–10 cancer at onset—intercept (θ_0)	0.0501
Probability of Gleason score 8–10 cancer at onset—linear term (θ_1)	0.0000
Probability of Gleason score 8–10 cancer at onset—quadratic term (θ_2)	0.0002
Hazard of metastasis—proportionality constant for PSA (γ)	0.00029
Hazard of clinical diagnosis (Localized, Gleason 2–7)—proportionality constant for PSA ($\chi_{L,2-7}$)	0.00288
Hazard of clinical diagnosis (Localized, Gleason 8–10)—proportionality constant for PSA ($\chi_{L,8-10}$)	0.00031
Hazard of clinical diagnosis (Metastatic, Gleason 2–7)—proportionality constant for PSA ($\chi_{M,2-7}$)	0.34852
Hazard of clinical diagnosis (Metastatic, Gleason 8–10)—proportionality constant for PSA ($\chi_{M,8-10}$)	0.03768

Figure 1.1 shows observed (black line) and model-predicted (gray line) incidence rates by age and stage for the calibration (1975–2000) and validation (2001–2010) years.

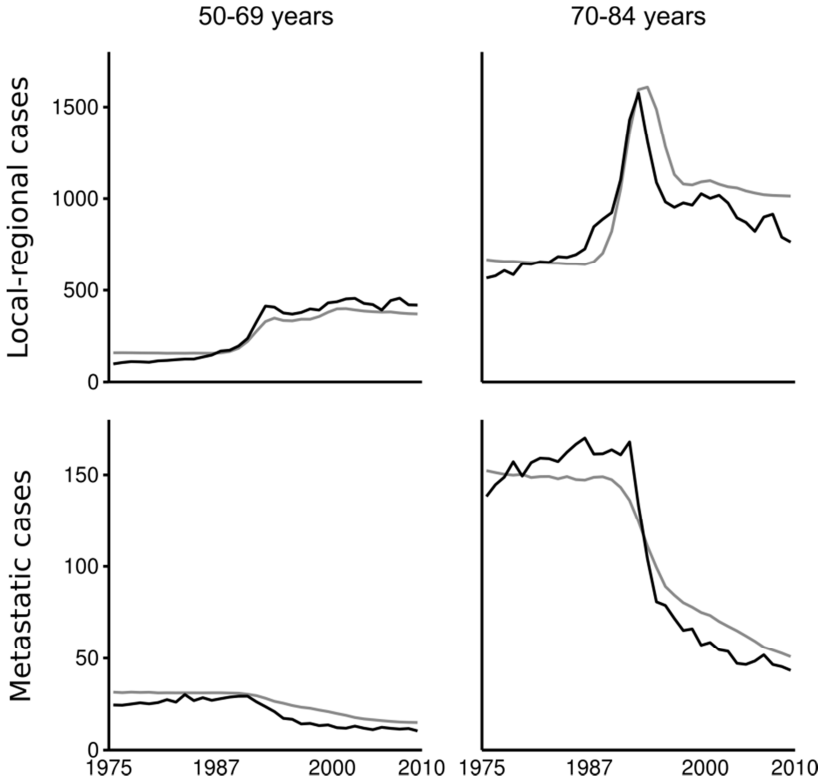

Figure 1.1. Observed and model-predicted incidence rates by age and stage.

1.3 Treatment and survival

1.3.1 Initial treatment

Men in the simulated population who are diagnosed can have either localized or metastatic cancer at diagnosis. Men who have metastatic cancer at diagnosis are assumed to receive palliative care that does not affect cancer-specific survival (described below). Men who have localized cancer at diagnosis are randomly assigned initial treatment based on frequencies of treatments (surgery, radiation, or conservative management) [from SEER] with or without adjuvant hormones [from the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE)] by age, year, stage, and grade at diagnosis. The treatment frequencies are frozen in 2010, the most recent year SEER data are available, so that men diagnosed with localized cancer after 2010 are assigned initial treatments using the 2010 frequencies.

1.3.2 Prostate cancer survival and treatment benefit after clinical presentation

Men who present clinically and who are assigned conservative management or hormone monotherapy have cancer-specific survival similar to that for untreated patients observed in SEER in 1983–1986, just before the first wave of PSA screening in the US. Men who are assigned surgery or radiation with adjuvant hormones receive the same survival improved by a hazard ratio $\lambda = 0.62$ (Bill-Axelson et al., 2011). Men who are assigned radiation without hormones receive survival improved by a hazard ratio $\lambda = 0.9$ before 1990, by a hazard ratio on the straight line between $\lambda = 0.9$ and $\lambda = 0.7$ in the period 1990–1995, and by a hazard ratio $\lambda = 0.7$ after 1995 (Etzioni et al., 2012). Formally, cancer-specific survival after clinical presentation is:

$$T_c \sim S(t \mid a_c, z_c)^{\lambda_c}$$

where a_c is patient age, z_c is the stage and grade of the tumor at age a_c, and λ_c is the hazard ratio representing benefit of initial treatment. For example, a 66-year-old man who is diagnosed with

localized cancer with Gleason score 2–7 and who receives surgery has cause-specific survival T_c randomly drawn from the adjusted SEER survival curve $S(t \mid 66, \text{localized}, \text{Gleason } 2-7)^{0.62}$.

1.3.3 Overdiagnosis and benefit of early detection

There are two types of screen detections: (1) early detections, men who would have presented clinically in the absence of screening, and (2) overdiagnoses, men who would *not* have presented clinically in the absence of screening. By definition, an overdiagnosed man cannot die of prostate cancer, so his cancer-specific survival is 1. Also by definition, an early detected man cannot die before his would-be point of clinical presentation, so a traditional "stage-shift" assumption implies his cancer-specific survival is:

$$T_s \sim \begin{cases} 1 & a_s \le t \le a_c \\ S(t \mid a_s, z_s)^{\lambda_s} & a_c < t \end{cases}$$

where a_s is his age at screen detection ($a_s \le a_c$), z_s is the stage and grade of the tumor at age a_s, and λ_s is the hazard ratio representing benefit of initial treatment, which is generated independently from initial treatment at clinical presentation. This is called a "stage-shift" assumption because a man diagnosed early by screening has survival corresponding to his age and tumor characteristics at screen detection, where much of the benefit comes from detecting a would-be metastatic cancer while it is still localized and treatable.

However, there are three potentially unsatisfying aspects of the traditional "stage-shift" assumption. First, survival drops to a lower level at onset of metastasis rather than representing a more continuous deterioration. Second, it seems more clinically realistic for the benefit of early detection to depend on the earliness of detection, e.g., for cancers that are detected much earlier to confer greater benefit than for cancers detected closer to the would-be point of clinical presentation. Third, and most importantly, the benefit of screening is rigid and cannot be calibrated to data (Wever et al., 2011).

These three criticisms can be overcome using a parameterized generalization of the traditional stage-shift model. In the FHCRC model, an early detected man has post-lead-time cancer-specific survival derived from a weighted average of survival times after stage-shifted screen detection and after clinical presentation, with greater weight placed on survival after stage-shifted screen detection for men with longer lead times (i.e., with longer $\ell = a_c - a_s$). Formally, an early detected man has post-lead-time cancer-specific survival:

$$T_s' = (1-w) \times T_s + w \times T_c$$

where the weight $w = \exp(-\alpha \times \ell)$ decreases exponentially with lead time at rate α. When the lead time ℓ is 0 (i.e., when $a_s = a_c$), the weight w is 1 and survival time is T_c, i.e., there is no early detection so no cancer-specific survival benefit. As the lead time ℓ increases, the weight w shrinks, and cancer-specific survival time approaches T_s (i.e., the full "stage-shift" applies). The rate α controls the rate at which the longer lead time translates into greater benefit, and this parameter can be calibrated to an observed dataset.

1.3.4 Estimation

To estimate the exponential rate α governing the benefit of early detection, we simulated the European Randomized Study of Screening for Prostate Cancer (ERSPC) through 11 years of follow-up. Specifically, we simulated a population of men ages 55–69 years screened every 4 years up to age 74 with biopsy referral for PSA > 3.0 ng/mL; this is an approximation to the protocol used in most centers. A random

86% of men referred to biopsy received biopsy, and a random 80% of men with latent cancer who received biopsy were diagnosed. We also simulated an identical population under no screening. Given total prostate cancer deaths in the two populations after 11 years of follow-up, we used univariate grid search to find a value of α that reproduced the 29% relative mortality reductions due to screening reported by ERSPC investigators after correction for non-compliance (Schröder et al., 2012). The same approach was used to find a value of α that reproduced a hypothetical 15% relative mortality reduction in the same setting. Results are shown in Table 1.3.

Table 1.3. Results of early detection model estimation

Parameter	Estimate
Exponential rate of benefit to reproduce 29% mortality reduction ($\alpha_{0.29}$)	0.550
Exponential rate of benefit to reproduce 15% mortality reduction ($\alpha_{0.15}$)	0.068

References

Gulati R, Inoue L, Katcher J, Hazelton W, Etzioni R. Calibrating disease progression models using population data: a critical precursor to policy development in cancer control. *Biostatistics.* Oct 2010;11(4):707-719.

Gulati R, Gore JL, Etzioni R. Comparative effectiveness of alternative prostate-specific antigen-based prostate cancer screening strategies: Model estimates of potential benefits and harms. *Ann. Intern. Med.* 2013;158(3):145-153.

National Center for Health Statistics. Vital statistics of the United States, Volume II: Mortality, part A. Washington DC: Government Printing Office; various years.

Mariotto A, Etzioni R, Krapcho M, Feuer EJ. Reconstructing prostate-specific antigen (PSA) testing patterns among black and white men in the US from Medicare claims and the National Health Interview Survey. *Cancer.* 2007;109(9):1877-1886.

Pinsky PF, Andriole GL, Kramer BS, Hayes RB, Prorok PC, Gohagan JK. Prostate biopsy following a positive screen in the Prostate, Lung, Colorectal and Ovarian cancer screening trial. *J. Urol.* Mar 2005;173(3):746-750; discussion 750-751.

Bill-Axelson A, Holmberg L, Ruutu M, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. *N. Engl. J. Med.* May 5 2011;364(18):1708-1717.

Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: Treatment changes and prostate cancer mortality declines. *Cancer.* May 17 2012.

Wever EM, Draisma G, Heijnsdijk EA, de Koning HJ. How does early detection by screening affect disease progression? Modeling estimated benefits in prostate cancer screening. *Med Decis Making*. 2011 Jul-Aug;31(4):550-8.

Schröder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. *N. Engl. J. Med.* Mar 15 2012;366(11):981-990.

2. UMICH model

The UMICH model of prostate cancer natural history is a statistical model built by analytic mathematics rather than computer simulation algorithms. It consists of a set of integrated submodels that fold into a joint model of cancer incidence, disease presentation (stage and grade) at diagnosis, disease progression in screen-detected persons incorporating early detection and treatment benefit, and survival and cancer mortality. The system of submodels is designed so they can be fit sequentially to cancer incidence and survival data from SEER.

2.1 Marginal incidence model

2.1.1. The i-model

Suppose $A = \{a_1, a_2, \dots\}$ are random ages when the subject is screened (a screening schedule). We have used a two-stage model for the screening schedule point process A. Let $\lambda_{1S}(a,t)$ be the hazard of the first prostate-specific antigen (PSA) test for a man of age a in year t. Then the probability that a man born in year b will not be tested by age a is

$$G_{1S}(a \mid b) = \exp\left\{-\int_0^a \lambda_{1S}(\xi, b+\xi)d\xi\right\}.$$

We assume that, in men who already had their first PSA tests, secondary tests $\{a_2, a_3, \dots\}$ form a non-homogeneous Poisson process with intensity $\lambda_{2S}(a,t)$. Both intensities of PSA testing λ_{1S} and λ_{2S} are treated as known bivariate functions estimated by approximating the output of a random schedule generator developed by the National Cancer Institute (Mariotto et al., 2007).

The survival function $G_{SDx}(\xi \mid b, \mathrm{y})$ representing the probability of no screen diagnosis for a subject born in year b, with tumor onset at age y, and delay time since onset ξ is derived in Tsodikov et al. (2006) as

$$G_{SDx}(\xi \mid b, \mathrm{y}) = G_{1S}(\mathrm{y}+\xi \mid b) + [1 - G_{1S}(\mathrm{y} \mid b)]G_{2SDx}(\xi \mid b, \mathrm{y}, \mathrm{y}) +$$

$$\int_0^\xi [1-\alpha(\zeta)]f_{1S}(\mathrm{y}+\zeta \mid b)G_{2SDx}(\xi-\zeta \mid b, \mathrm{y}+\zeta, \mathrm{y})d\zeta,$$

where α is age-dependent screening sensitivity, f_{1S} is the probability density function (p.d.f.) corresponding to λ_{1S}, and

$$G_{2SDx}(\xi \mid b, a, \mathrm{y}) = \exp\left\{-\int_{\max(A_O-a,0)}^{\xi} \lambda_{2S}(a+\zeta, b+a+\zeta)\alpha(\zeta+a-\mathrm{y})d\zeta\right\},$$

and $\int_a^b = 0$ for any $b \le a$. The above expressions are a result of averaging over A and the Bernoulli outcomes of screening tests with probability of success α within the subject, given y. Using conditional independence of the competing risks of cancer diagnosis by screening and clinically, given age at tumor onset y, we have

$$\lambda_I(a \mid b) = -\frac{d}{da}\log E\,\{G_{CDx}(a-\mathrm{y} \mid b, \mathrm{y})G_{SDx}(a-\mathrm{y} \mid b, \mathrm{y})\},$$

where G_{CDx} is the survival function of the sojourn time.

The marginal incidence model (i-model) describes the risk to be diagnosed with prostate cancer at age a_I either clinically (CDx) or by PSA screening tests (SDx). The prostate cancer incidence for x birth cohort

(a cohort of men turning 50 in year x) can be written as a complex mixture model where missing data include the age at onset $A_O = y$, the screening schedule, and the detection process. The p.d.f. of cancer diagnosis at the age a_I given birth year x can be written as

$$f_I(a_I|x) = \int_0^{a_I} f_o(y|x) f_I(a_I - y|x, y)\, \mathrm{d}y,$$

where f_o is the unconditional p.d.f. of age at tumor onset and $f_I(a_I - y|x, y)$ is the p.d.f. of age at cancer diagnosis given birth year x and tumor onset time y. Under competing risks, $f_I(a_I - y|x, y)$ can be split into two crude densities,

$$f_I(a_I - y|x, y) = f^c_{CDx}(a_I - y|x, y) + f^c_{SDx}(a_I - y|x, y),$$

where

$$f^c_{CDx}(a_I - y|x, y) = f_{CDx}(a_I - y|x, y) G_{SDx}(a_I - y|x, y)$$

and

$$f^c_{SDx}(a_I - y|x, y) = f_{SDx}(a_I - y|x, y) G_{CDx}(a_I - y|x, y).$$

Here $f_{CDx}(a_I - y|x, y)$ and $G_{CDx}(a_I - y|x, y)$ are the p.d.f. and survival function (s.f.) of T_{CDx},, the sojourn time distribution, and $f_{SDx}(a_I - y|x, y)$ and $G_{SDx}(a_I - y|x, y)$ are the p.d.f. and s.f. of T_{SDx}, the time to PSA diagnosis, respectively, computed as an average over the screening schedule point process and the outcomes of screening tests. Integrating out the age at tumor onset y, we have the unconditional p.d.f. of age at diagnosis a_I

$$f_I(a_I|x) = f^c_{CDx}(a_I|x) + f^c_{SDx}(a_I|x).$$

Note that the unconditional p.d.f. $f_{CDx}(a_I|x)$ is a function of the age of the subject while the conditional p.d.f. $f_{CDx}(a_I - y|x, y)$ is a function of the delay time defined as $t_D = a_I - y$.

2.1.3. Estimation

Maximum likelihood is used to fit the model. Conditional on the birth year x, the likelihood is defined as one for a survival model where age at diagnosis serves as the survival time and there is a cross-sectional observation structure in which data represent observations of the risk set (SEER male population ages 50–84 in years 1975–2000) and cases of cancer diagnosis serve as events:

$$l = \sum_a \sum_t P(a, t)\{\lambda_{Obs}(a|x)\log(\lambda_I(a|x)da) - \lambda_I(a|x)da\}.$$

Here a is age; $P(a, t)$ is population count with age a and calendar time $t = a + x$; $\lambda_{Obs} = C(a, t)/P(a, t)$ is the observed incidence with $C(a, t)$ being the count of cancer diagnoses; and da is 1 year. The likelihood is maximized by the method of conjugate gradients. Details and estimates of the model parameters can be found in Tsodikov et al. (2006).

2.1.4 Prediction of the lead time

Cancer development passes through the disease-free state and the pre-clinical state before being detected or censored without a diagnosis. The subject's age at tumor onset A_O represents the duration

of the disease-free state. The duration of latent cancer growth in the pre-clinical state is given by the delay time $\xi_D = A - A_O$, a backward recurrence time, represented by the period between cancer onset at age A_O and its diagnosis at age A. Given onset time, cancer diagnosis is a result of two competing risks, the potential time to detection by screening, ξ_{SDx}, and the potential time to clinical diagnosis due to symptoms of the disease, ξ_{CDx}, so that the delay time is $\xi_D = \min(\xi_{SDx}, \xi_{CDx})$. The time ξ_{CDx} is referred to as the sojourn time. The forward recurrence time

$$\xi_L = \begin{cases} \xi_{CDx} - \xi_{SDx}, & \xi_{CDx} > \xi_{SDx}, \\ 0, & \xi_{CDx} \le \xi_{SDx}, \end{cases} \quad \begin{matrix} \text{Screening diagnosis } (SDx) \\ \text{Clinical diagnosis } (CDx) \end{matrix}$$

is called the lead time (as adopted here—multiple definitions are available in the literature), and it describes how much diagnosis was advanced due to screening. Note that the lead time is zero in patients detected clinically. The joint distribution of the lead-time (w argument) and age at diagnosis A (a argument) is given by an average over age at cancer onset A_O (y argument) as

$$f_{LT}(w, a|x) = \int_0^a f_O(y|x) f_{CDx}(a - y + w|x, y) \begin{cases} G_{SDx}(a - y|x, y), & w = 0 \\ f_{SDx}(a - y|x, y), & w > 0 \end{cases} dy, \quad (1)$$

where f_O is a p.d.f. of the age at onset, and f_{CDx} and f_{SDx} are p.d.f.s and G_{CDx} and G_{SDx} are s.f.s for the two competing risks of CDx and SDx (corresponding to random variables ξ_{CDx} and ξ_{SDx}), respectively. The expression under the integral is a joint p.d.f./p.m.f. $f_{LT}(w, a, y|x)$ of age at onset $A_O = y$, diagnosis (any type) at age $A = a$, and potential clinical diagnosis at $A + \xi_L = a + w$, if applicable (in case real diagnosis is SDx). Note that zero lead time ($w = 0$ in (1)) corresponds to observed CDx occurring at age $A = a$ and that f_{LT} has a mass at this point corresponding to clinical diagnosis (sometimes referred to as *interval* detection to emphasize that CDx occurred in an interval between screens). The conditional lead-time distribution is obtained from the joint one as

$$f_{LT}(w|a, x) = \frac{f_{LT}(w, a|x)}{f_I(a|x)},$$

where f_I is the p.d.f. corresponding to the marginal cancer incidence rate λ_I. The conditional age at tumor onset $[Y|a, x]$ serves as a mixed effect in the model of disease presentation at diagnosis discussed in the next section.

2.2. Disease presentation (stage, grade) at diagnosis

2.2.1. The z-model

Disease stage is dichotomized into the local-regional (LR) stage and distant (D) stage. Disease grade is dichotomized into well or moderately (WM) differentiated (low grade) and poorly differentiated or undifferentiated (PU) disease (high grade). We use the random variable Z to denote four possible combinations of binary stage and grade classifications (Localized/Reginal Stage, Low Grade)=LL, (Localized/Reginal Stage, High Grade)=LH, (Distant Stage, Low Grade)=DL, (Distant Stage, High Grade)=DH. Generally, Z can be any multivariate mark on the cancer incidence process. The Z-specific incidence model describes the probability of being diagnosed with prostate cancer at a certain age and with specific stage and grade z. Conditional on birth year x, age at tumor onset y, and age at tumor

diagnosis a_I, the probability of being diagnosed with stage and grade z was modeled using a mixed multinomial logit model (Chefo and Tsodikov, 2009). Missing data include the delay time between tumor onset and diagnosis (a backward recurrence time) and the mode of diagnosis. Calendar time and age are treated as fixed effects covariates. Using the Z-specific incidence model, we predict the conditional (multinomial) distribution of z, $f_I(z|x,a_I)$ given the cohort x and incident age a_I. Note that $f_I(z|x,a_I)$ is a conditional average over tumor onset age Y and mode of diagnosis

$$I_{SDx} = \begin{cases} 1, & if\ \ SDx \\ 0, & if\ \ CDx \end{cases}.$$

Using the model we can update the distribution of Y given the information available at diagnosis.

A mixed multinomial model is used for the conditional distribution of stage and grade at diagnosis, given age at onset, $f(z|a,x,y) = \Pr\{z|A=a,\ \text{Birth year } = x, \text{Age at onset} = \text{y}\}$. On taking an average with respect to missing age at onset (or lead time), $f(z|a,x) = E\,\{f(z|a,x,\text{Y})|a,x\}$, we obtain the factor partitioning the marginal incidence λ_I into the z-specific components

$$f_I(z|x,a) = \frac{f_I(a,z|x)}{f_I(a|x)} = \frac{\int_0^a f_I(z|x,a,y) f_I(a|x,y) f_o(y|x)\mathrm{d}y}{f_I(a|x)},$$

$$\lambda_I(a|x) = \sum_z \lambda_I(a,z|x), \quad \lambda_I(a,z|x) = \lambda_I(a|x) \times f(z|a,x),$$

where the expectation is taken over the conditional distribution of age at cancer onset, $Y|a,x$, given age at diagnosis and birth year.

2.2.2. Estimation

Maximum likelihood is used to fit the model. Conditional on the birth year x, the likelihood is defined as one for a *marked* survival model where (a) age at diagnosis serves as the survival time and (b) the combination of stage and grade serves as a mark. The conditional likelihood used to estimate the submodel for the mark uses age at diagnosis as a covariate and is based on observations of the cases only at cancer diagnosis:

$$l = \sum_a \sum_t \sum_z C(a,t)\log\big(f(z|a,x)\big), \quad t = a + x.$$

Here a is age, $t = a + x$ is calendar time, and $C(a,t)$ is the count of cancer diagnoses. The likelihood is maximized by the method of conjugate gradients. Details and estimates of the model parameters can be found in Chefo and Tsodikov (2009).

2.2.3. Updated prediction of the lead time

The conditional distribution of age at tumor onset Y can be written as

$$Y \sim f_I(y|x,a_I,z) = \frac{f_I(a_I,z,y|x)}{f_I(a_I,z|x)} = \frac{f_I(z|x,a_I,y) f_I(a_I|x,y) f_o(y|x)}{\int_0^{a_I} f_I(z|x,a_I,y) f_I(a_I|x,y) f_o(y|x)\ \mathrm{d}y}.$$

The lead time distribution can then be updated with the additional information on the disease presentation index z observed at diagnosis as

$$f_{LT}(w|a,z,x) = \frac{f_{LT}(w,a,z|x)}{f_I(a|x)}, \quad f_{LT}(w,a,z|x) = \int_0^a f_{LT}(w,a,y|x) f(z|a,x,y)\, dy.$$

Details and estimates are available in Chefo and Tsodikov (2009). The lead time in distant stage disease is small as cancers are growing faster and are typically missed by screening. Lead time is largest for men around 70 years of age when most prostate cancers are detected. The conditional updated age at tumor onset $[Y|a,z,x]$ serves as a mixed effect in the model of disease progression discussed in the next section. Shown in Figure 2.1 are model-based predictions of the mean lead time by the characteristics observed at diagnosis.

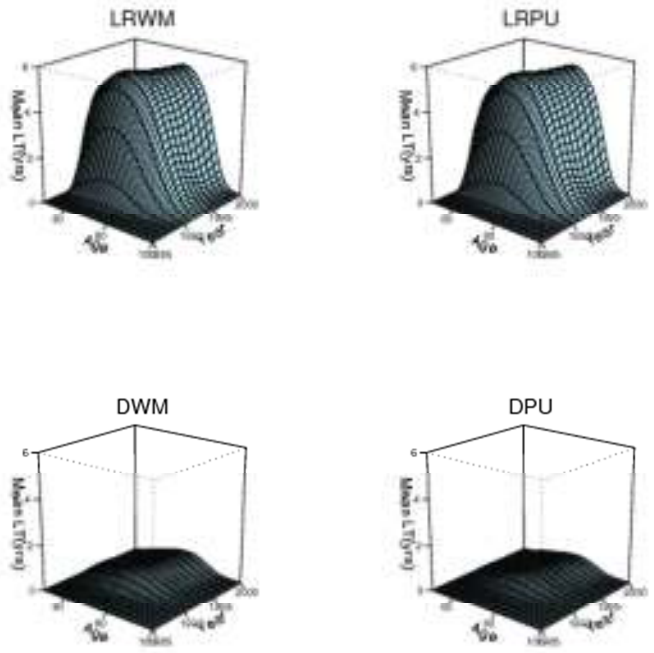

Figure 2.1. Model-predicted mean lead time by stage, grade, age, and year of diagnosis.

Note that the i- and z- submodels used in this study were previously described in Tsodikov et al. (2006) and Chefo and Tsodikov (2009). The p- submodel, described in the next section, represents an original extension to reflect disease progression after screen detection and benefits of early detection and primary treatment.

2.3. Disease progression model post screen diagnosis

2.3.1. The p-model

Let $Z(\xi)$ be the cancer progression process with the time (age of tumor) ξ counted out from the point of cancer onset in the subject. Given the two potential competing risks of clinical (CDx) and screening (SDx) diagnosis, we can define the corresponding potential values of the cancer development process $Z(\xi_{SDx})$ and $Z(\xi_{CDx})$ measured on the same subject. The competing character of the two detection mechanisms

makes them partially unobserved. At most we can observe age at diagnosis $A = Y + \min(\xi_{SDx}, \xi_{CDx})$ and the disease characteristics $Z(\xi_{SDx})$ if the observed diagnosis is by screening and $Z(\xi_{CDx})$ otherwise, but not both. Note that since ξ_{SDx} is undefined for an *unscreened* subject, we would not be able to treat ξ_{SDx} as missing data for the subject to define his likelihood contribution. Consequently, we would not be able to use a likelihood-based approach to fit the joint model. Let the indicator

$$I_{SDx} = \begin{cases} 1, & SDx \\ 0, & CDx \end{cases} = I(\xi_L = 0)$$

express the type of observed diagnosis. Let the vector $V = (a, z)$ be the disease presentation at diagnosis combining age and the disease severity z at the point of diagnosis. Conditional on the specific type of diagnosis, we have $V_1 = (Y + \xi_{SDx}, Z(\xi_{SDx}))$ given SDx and $V_0 = (Y + \xi_{CDx}, Z(\xi_{CDx}))$ given CDx. For a screen-detected patient, V_1 is observed and V_0 is the unobserved hypothesized disease presentation at potential CDx. The stage and grade specific incidence model (z-model) specifies the distribution of $V = V_{I_{SDx}}$ for diagnosis of any type (I_{SDx} is random) and explains its variance by the heterogeneity of the disease natural history and the screening process preceding diagnosis.

The disease progression model (p-model) defines the probability of disease progression during the lead-time in the absence of treatment represented by the transition model $[V_0 \mid V_1]$. Jointly with the lead-time distribution, the disease progression model can provide a joint distribution of (V_1, V_0) for screen - detected patients. However, V_1 is observed on one set of patients and V_0 on another, independent set of patients. With population data, I_{SDx} is also unobserved, so we observe marginal V on patients diagnosed in the screening era and V_0 before screening was introduced in the population. This observation structure precludes the traditional approach of fitting a transition model $[V_0|V_1]$ to longitudinal data V_1, V_0. Nevertheless, the way screening affects the marginal V has information allowing one to identify the transition model. For a screen-detected subject, let $f_V(V_0|V_1, x)$ be the joint p.d.f./p.m.f. of the disease presentation at potential CDx (V_0) conditional on the observed presentation V_1 at SDx and the birth cohort x. Let the baseline (under the null hypothesis of no interaction between screening and treatment, STx) probability $p_b(z_0|z_1, \xi_L)$ describe the the transition probability for the disease severity index z between SDx and potential CDx, given the lead time ξ_L. Then we can expand f_V as

$$f_V(V_0|V_1, x) = f_{LT}(w|a_1, z_1, x, \xi_L > 0) p_b(z_0|z_1, \xi_L = w),$$

where

$$f_{LT}(w|\cdot, \xi_L > 0) = \frac{f_{LT}(w|\cdot)}{1 - f_{LT}(0|\cdot)}$$

is conditional on SDx that is equivalent to a positive lead time, $\xi_L > 0$. While generally p_b may depend on the lead time and other covariates, in the data analysis of this paper we assume they are a set of unknown parameters arranged in the form of the so-called progression probability matrix (PPM) with i, j elements $p_{bij} = p_b(z_0 = j|z_1 = i)$, where $i \le j$, $i, j = 1, \ldots, 4$ index the four categories of stage and grade z. The fact that $p_{bij} = 0$ for $i > j$ reflects the assumption that cancer cannot regress. The PPM is

shown in Table 2.1. Under the null hypothesis of no interaction between screening and treatment, the baseline PPM probabilities p_b are not affected by treatment applied at the point of SDx. Now consider two model predictions:

1. z-specific incidence $\lambda_I(a, z|\neg S)$ under no screening ($\neg S$, zero screening sensitivity); and

2. the model predicted z-specific incidence $\lambda_I(a, z|IS)$ under ignored screening (IS), the patient is left undiagnosed until the end of his lead time.

While the first scenario does not involve the PPM and is expressed by the z-specific incidence model prediction under zero screening sensitivity, the second scenario uses PPM to predict stage and grade at the end of the lead time.

The IS counterfactual scenario operates under the null hypothesis of no interaction between screening and treatment (or no benefit of screening regarding cancer progression during the lead time). Then the progression model must satisfy

$$\lambda_I(a, z|\neg S, x) \equiv \lambda_I(a, z|IS, x), \qquad (2)$$

where $\equiv$ denotes a uniform equality over a, z, x.

Distributional characteristics of z-marked cancer incidence under IS can be deduced from the following joint p.d.f.:

$$f(a, z|IS) = f(a, z, SDx|IS) + f(a, z, CDx|IS)$$

$$= \int_0^a \sum_{z_1 \le z} f(a_1, z_1|S) f_{LT}(a - a_1|z_1, a_1) p_b(z|z_1, a - a_1) \mathrm{d}a_1$$

$$+ f(a, z, CDx|S) \times f_{LT}(0|a, z).$$

This expression represents an incident cancer under IS as either an observed CDx with zero lead time (p.d.f. $f(a, z, CDx|IS)$, second term) or an observed SDx (first term) with a potential CDx later, in which case possible presentations at SDx prior to the potential CDx are entertained in the first term of the sum. The integration over a_1 and summation over z_1 represent a formula of total probability exercised over natural history of the disease process preceding the occurrence of symptoms at age a and stage z. Note that the clinical incidence is the same under the actual screening scenario (S) and under the hypothetical IS, $f(a, z, CDx|S) \equiv f(a, z, CDx|IS)$.

The progression mechanism under screening (entering the equality (2) through $\lambda_I(a, z|IS)$ that is computed using the baseline probabilities p_b) characterizes the null hypothesis of treatment effect on cancer progression. In case of more general progression models, not necessarily formulated in terms of PPM, this represents the null hypothesis expressing the general equivalence of ignored screening (IS), zero screening sensitivity, and zero treatment effect. In other words, there is no difference between no screening, screening with zero sensitivity, or screening combined with ineffective treatment regarding the disease presentation V_0 at the point of symptoms (real or counterfactual CDx).

Note that the overall null hypothesis of the combined benefit of screening and treatment involves a consideration of survival in addition to cancer progression. An equality similar to (2) for cancer mortality can be defined in addition to the cancer progression part or, alternatively, an equality for survival times post the guaranteed lead-time survival. The null hypothesis is directly relevant to the so-called active surveillance or watchful waiting regimen when aggressive treatment is deferred until signs of progression emerge. Assuming this happens at the point of occurrence of symptoms (CDx), the model under IS approximately describes the evolution of a screen-detected patient under such conservative management. In this case, treatment may be effective and, applied at CDx, may improve post lead-time survival. However, being deferred, it misses the opportunity to affect latent cancer progression (PPM) before CDx. Early detection by screening is not a modifier to the treatment effect as far as cancer progression goes.

2.3.2. Estimation

Motivated by the above discussion, to estimate the PPM we treat the equality of counterfactual hazards as an equation for the unknown p_b patrameters. The idea is to deduce the disease progression model from the marginal one by making the two predictions on both sides of (2) as close as possible. The target function to be minimized for the estimate of the disease progression model, l, can be written as the Poisson likelihood "distance" between the two predictions. The main difficulty in estimating PPM is rooted in the fact that CDx and SDx are not observed on the same subject. So there is no direct subject-specific data on the disease progression. To estimate the PPM, we notice the equality of the model predictions in the following two counterfactual scenarios: (1) under no screening ($\neg S$, zero screening sensitivity) and (2) the model prediction as if screening were ignored (IS) and the patient was left undiagnosed until the end of his lead time. While the first scenario does not involve the PPM, the second counterfactual scenario uses PPM to predict stage and grade at the end of lead time. The idea is then to deduce the disease progression model from the marginal one by making the two counterfactual predictions as close as possible. The target function to be minimized for the estimate of the disease progression model, l, can be written as the Poisson likelihood "distance" between the two predictions treating one of them as "observed" data ($\neg S$) and the other as expected (IS).

Let a_i and $z_i, i = 0,1$, be age and stage/grade at CDx and SDx, respectively. Conditional on the birth year, x, define a target function to be maximized as

$$l = \sum_{a_0} \sum_t \sum_{z_0} P(a_0, t, z_0)\{\lambda_o(a_0, z_0|\neg S)\log\lambda_e(a_0, z_0|iS) - \lambda_e(a_0, z_0|iS)\},$$

where $P(a_0, t, z_0)$ is the population count with age a_0, stage and grade z_0 in calendar year t, and λ_e is deduced from the corresponding joint p.d.f.

$$f(a_0, z_0|iS) = f(a_0, z_0, SDx|iS) + f(a_0, z_0, CDx|iS)$$

$$= \int_0^a \sum_{z_1 \le z_0} f(a_1, z_1|S) f_{LT}(a_0 - a_1|z_1, a_1) p(z_0|z_1, a_0 - a_1)\mathrm{d}a_1$$

$$+ f(a_0, z_0|S) \times f_{LT}(0|a_0, z_0).$$

Standard errors of the estimates are obtained by bootstrap.

Table 2.1. Results of the Estimated Progression Probability Matrix (PPM)

		Counterfactual Clinical Diagnosis			
Screening Diagnosis	Stage/Grade	LR/L (1)	LR/H (2)	D/L (3)	D/H (4)
	LR/L	p_{b11}=0.893	p_{b12}=0.051	p_{b13}=0.056	p_{b14}=0
		(0.879,0.906)	(0.039,0.062)	(0.049,0.064)	(0,0)
	LR/H	0	p_{b22}=0.717	0	p_{b24}=0.283
			(0.676, 0.757)		(0.243, 0.324)
	D/L	0	0	p_{b33}=1	p_{b34}=0
				(1,1)	(0,0)
	D/H	0	0	0	1
Stage: LR=Local/Regional, D=Distant. Grade: L=Low (WM), H=High (PU). 95% confidence intervals are in parentheses					

Figure 2.2 shows observed (black line) and model-predicted (gray line) incidence rates by age and stage for the calibration (1975–2000) and validation (2001–2010) years.

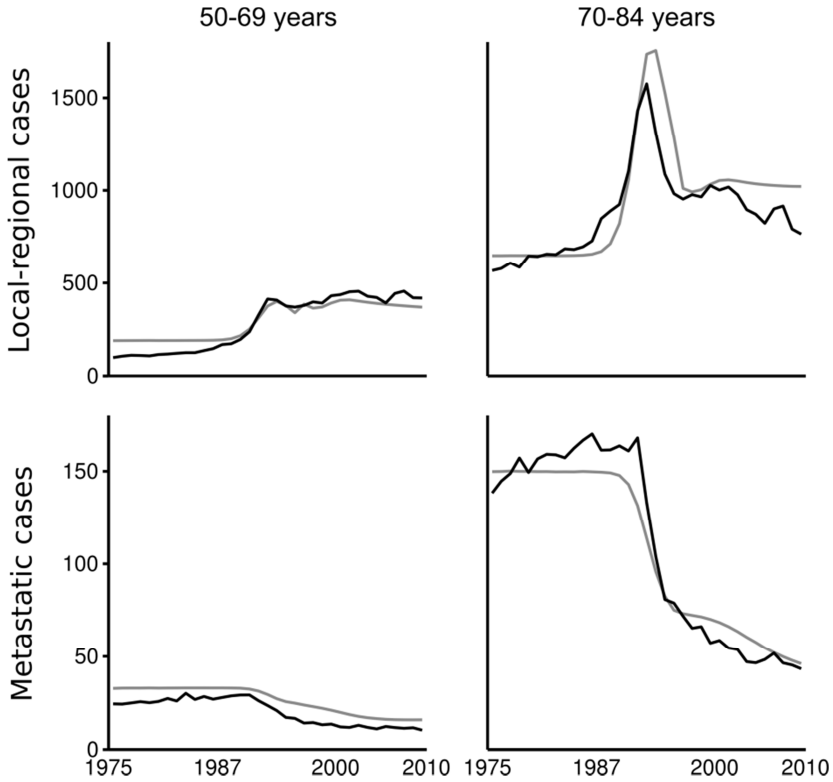

Figure 2.2. Observed and model-predicted incidence rates by age and stage.

The progression model was fit to SEER data (Table 2.1). Only 5% to 6% of patients progress in stage/grade respectively in the localized stage, low grade group at SDx. No progression from the best category (local stage and low grade) to the worst category (distant stage and high grade) occurs, perhaps due to the long time frame required for such a big transition compared to the lead time. Localized stage, high -grade patients are more likely to progress. About 28% potentially progress to distant stage. There is no grade progression in the distant stage patients likely because their lead time is too short for the grade to change.

2.3.3. Screening benefit. Treatment effect on disease progression after screen diagnosis.

If treatment had an effect, the baseline probabilities p_b would be transformed by a categorical regression model with treatment as a covariate and p_b corresponding to the baseline of no treatment effect. Consider a cumulative logit regression (proportional odds, PO) model working with a row of the PPM matrix corresponding to certain observed disease characteristics at SDx. Let η_{Tx} be the cumulative odds ratio expressing the difference between the conditional distribution of $z_0|z_1$ with vs without treatment, the latter given by $p_b(\mathrm{z}_0|z_1)$. Under the PO, model the transition probability under the treatment effect is

$$p_{ij} = \Pr(Z_0 = j|Z_1 = i, Tx) = \frac{p_{bij}\eta_{Tx}}{p_{bij}(\eta_{Tx} - 1) + 1}, j > i,$$

where $z_{0,1}$ is stage-grade response at CDx and SDx, respectively.

In the cumulative logit model, let cp_b be the cumulative baseline transition probability. The cumulative transition probability given the treatment effect is

$$cp_{ij} = \Pr(z_0 \le j|z_1 = i, Tx) = \frac{cp_{bij}/\eta_{Tx}}{cp_{bij}(1/\eta_{Tx} - 1) + 1}, j \ge i,$$

and $p_{ij} = cp_{ij} - cp_{i(j-1)}$.

When there is no treatment effect on the probabilities (i.e., $\eta_{Tx} = 1$), $p_{ij} = p_{bij}$. When the treatment effect is really large (i.e., $\eta_{Tx} \to 0$), the PPM will become an identity matrix corresponding to the full stage-shift when the stage and grade carry over from the screen diagnosis to the counterfactual clinical diagnosis.

In the extreme, treatment applied at the point of SDx may completely arrest cancer progression, in which case PPM would be an identity matrix corresponding to the full stage-shift assumption.

The treatment model describes the probability of receiving a certain treatment combination at the time of cancer diagnosis. Using SEER data, we classified treatments into three major categories: conservative management (CM), radiation therapy (RT), and radical prostatectomy (RP). Hormone therapy (HT) is commonly used as an adjuvant therapy following one of those three primary treatments, RT most commonly. Note that HT information is not available in cancer registries, hence we used a two-stage model to predict treatment allocations. In the first stage, we modeled the probability of receiving one of the three treatments defined in SEER data conditional on birth year x, age at diagnosis a_I, and grade

using multinomial logit model. In the second stage, we conditioned on the treatment obtained from the first stage, birth year x, age at diagnosis a_I, and grade and used logistic regression to model the probability of receiving the additional adjuvant hormone therapy.

By combining first and second stage, we can obtain the probability of actual treatment received given birth year x, age at diagnosis a_I, and grade. These model-based predicted probabilities and empirical SEER estimates are shown in Figure 2.3.

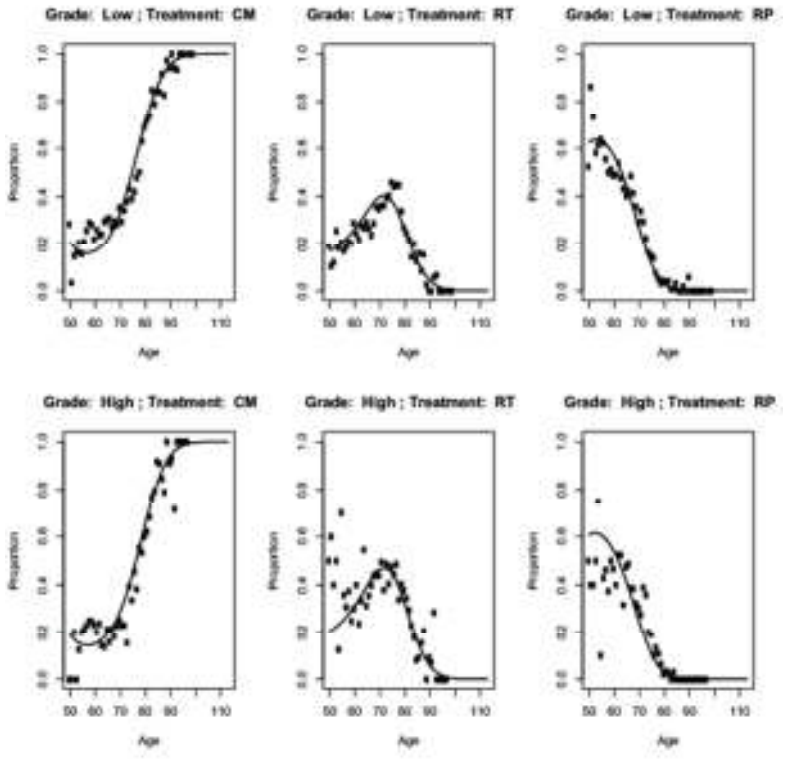

Figure 2.3. Observed and model-predicted probabilities of SEER treatment by age at diagnosis and grade in localized disease.

2.4. Survival and mortality

2.4.1. The s-model

The survival function G describes the time spent in the clinical state (from diagnosis to cancer-specific death in the absence of other causes) conditional on age at incidence a_I, year of diagnosis t ($t = x + a_I$), stage and grade Z, and treatment Tx.

The generalized stage-shift model is motivated by the traditional stage-shift model by allowing disease progression during the lead time. Two adjustments were made in the survival model during the PSA era. First, lead -time adjustment was made to make sure survival times are always measured from the time of clinical diagnosis to time of death, implying a guaranteed lead-time benefit. Additionally, survival is conditional on the stage and grade at the time of clinical diagnosis, the latter being unobserved if the patient is screen-detected. Three treatment effects were included in the model. The treatment main

effect θ_{Tx} describes how treatment affects post-lead-time survival, conditional on stage and grade at the end of the lead time. The treatment effect applied to the disease progression (η_{Tx}, estimated as 0.3 for the ERSPC-sized effect) measures how treatment prevents stage and grade progression during the lead time (an unobserved process in a screen-detected person). Finally, even if the patient progresses during the lead time, the one treated earlier could still be better off than the one treated close to the point of symptoms, and this is represented by the treatment-by-lead-time interaction effect, $\theta_{Tx\times LT}$, in the s.f. G post real or counterfactual CDx. In this paper, the latter effect is not used.

With all the above assumptions, the adjusted survival function for patients diagnosed in the screening era becomes

$$G(t_s|a_I,t,z,Tx) = G_{LT}(t_s|a_I,t,z) \quad \text{Guaranteed survival during lead time}$$

$$+\ f_{LT}(0|a_I,t,z)G_b(t_s|a_I,t,z,Tx) \quad \text{Survival after clinical diagnosis}$$

$$+\int_{0+}^{t_s}\sum_{z_0\geq z}\Pr(z_0|z,Tx)f_{LT}(s|a_I,t,z)G_b(t_s-s|a_I+s,t,z_0,Tx)\,\mathrm{d}s \quad \text{Survival after screen diagnosis}$$

where

$$G_b(t_s-s|a_I+s,t,z_0,Tx,s) = G_0(t_s|a_I,t,z_0)^{\theta_{Age}\theta_{Tx}\theta_{Tx\times LT}}$$

and G_0 is the model under a reference group under conservative management estimated using SEER survival before screening was introduced (1988).

2.4.1. Prediction of cancer mortality

Let A and Z be the age and tumor characteristics at the time of diagnosis, respectively. In the UMICH model, Z represents all four possible combinations of stage (SEER local-regional and distant) and grade (Gleason score 2–7 and 8–10).

Let $G(F \mid a, z)$ denote the survival function at F years of follow-up averaged over the distribution of treatment and adjusted for the lead time. This function combines lead time (W) and screening and treatment benefits. Cancer mortality is deduced from the cause-specific cancer survival probability at age A_M averaged over possible incident cancer:

$$G_M(a_M) = \mathrm{E}\left\{G(A_M - A \mid A, Z)\right\},$$

where expectation is taken over A and Z, and $G(F \mid A, Z) = 1$ when $F < 0$. Explicit expression of screening benefit requires that S be spelled out in detail through the s-model adjusted survival function G as $G(F \mid A,Z) = E[G(F \mid A,Z,\ Tx) \mid A,Z]$, averaged over the distribution of treatment applied for the patient diagnosed with characteristics Z at age A.

Conditional on birth year x, cancer mortality is a hazard function, λ_M, of the age at prostate cancer death, a_M. By definition,

$$\lambda_M(a_M|x) = \frac{f_M(a_M|x)}{S_M(a_M|x)},$$

where $f_M(a_M|x)$ is the p.d.f. and $S_M(a_M|x)$ is the s.f. averaged over the whole patient history from birth, including possible cancer incidence by screening or clinical diagnosis, disease characteristics at diagnosis, treatment applied, progression during the lead time (if screen-diagnosed), and survival post lead time. For a man in birth cohort x, the probability of prostate cancer-specific survival in the absence of other causes at age a_M can be written as

$$S_M(a_M|x) = \int_0^{a_M} f_I(a_I|x) \times \quad \text{Incident cancer at age } a_I$$

$$\sum_z f_I(z|x,a_I) \times \quad \text{Disease presentation with characteristics } z$$

$$\sum_{Tx} f_{Tx}(Tx|x,a_I,z) \times \quad \text{Treatment applied}$$

$$G(a_M - a_I|x,a_I,z,Tx)\mathrm{d}a_I + \quad \text{Lead-time adjusted post-diagnosis survival}$$

$$G_I(a_M|x). \quad \text{No incident cancer}$$

The parts of the equation describe the probability for a man who had prostate cancer diagnosed at age a_I (before age a_M) with stage and grade z, received treatment Tx at the time of diagnosis, and survived at least $a_M - a_I$ years after the diagnosis, and the probability for the man who has never been diagnosed with prostate cancer. Mortality is a convolution of five models addressing the development of the disease under screening and treatment interactions: Marginal incidence (i-model), Z-specific incidence (disease presentation at diagnosis, z-model), treatment, disease progression between screening and counterfactual clinical diagnosis (p-model), and survival model (s-model).

Mortality $\lambda_M(a_M|x)$ is deduced from $S_M(a_M|x)$ using the well-known relationship between hazard and survival functions in survival analysis:

$$\lambda_M(a_M|x) = -\frac{d}{da_M}\log S_M(a_M|x)\ .$$

References

Mariotto A, Etzioni R, Krapcho M, Feuer E. Reconstructing PSA testing patterns between black and white men in the US from Medicare claims and the National Health Interview Survey. Cancer 2007; 109(9):1877–1886.

Tsodikov A, Szabo A, Wegelin J. A population model of prostate cancer incidence. Statistics in Medicine 2006; 25(16):2846–2866.

Chefo S, Tsodikov A. Stage-specific cancer incidence: an artificially mixed multinomial logit model. Statistics in Medicine 2009; 28(15):2054–2076.