Supplemental Methods

1. Modeling philosophy
   a. Micro-simulation is an implementation of the Monte Carlo method, which is used to solve hard problems through rote computation as opposed to derivation. The primary challenge is to incorporate a wide variety of data types (distributions of life spans, diagnosis rates, survival rates, costs, screening test performance, etc.) from a variety of sources to solve a suite of problems related to effectiveness and cost-effectiveness of cancer screening. In our model, the various data types are divided into components consisting of related data and the structures of the data are modeled within each component.
   b. The components are used to generate the various aspects of a single woman's life history to whom screening is then applied. This process is repeated for a million women and the life experiences are tallied across all the simulated population.
   c. Screening outcomes are deterministic once the population (i.e. life histories of the cohort of one million women) has been generated. We can generate a single population and see how it responds to various screening scenarios, including no screening at all. In essence we are doing the ultimate controlled clinical trial, where we are allowed to observe alternate situations for each individual woman.

   We can observe each woman's survival both with and without screening (shown in Figure S1).

2. Natural History: For each member of the population we generate a series of life events which match the distributions of empirically observed data sources to build a series of life events for each woman in our population.

   a. Age of death: Each woman is assigned an age of death from other cause (i.e. not ovarian cancer) based on life tables for females in the United States .
   b. Age of clinical diagnosis of ovarian cancer: Individual women are assigned an age of clinical diagnosis of ovarian cancer based on observed incidence from the SEER registry. These women are characterized as ovarian cases. However the vast majority of women (>96%) who are not assigned an ovarian cancer diagnosis comprise the healthy and benign populations. Cases with an age of death from other cause prior to age of clinical diagnosis of ovarian cancer make up the latent ovarian cancer population.
   c. Tumor Characteristics: Women in the ovarian cancer case and latent populations are assigned tumor characteristics (stage, grade and histology) from a series of conditional distributions estimated using the SEER registry (see Table S5).
   d. Disease duration and stage length: Because Pre-clinical disease progression is generally not directly observable in humans we conducted a survey of physicians on their estimates of survival for patients who decline treatment as this portion of the disease history is potentially observable by doctors. To obtain the most reliable estimates possible we identified clinicians with ovarian cancer expertise. Accordingly, we amassed a small database of gynecological and medical oncologists from across the United States including individuals listed in all sections of the 2005 membership directory of Society of Gynecological Oncologists and faculty associated with gynecological oncology fellowship training programs. Physician names were compared against PubMed to identify physicians that had published on any topic related to ovarian cancer within 5 years of being invited to participate. Providers were also asked to categorize the number of ovarian cancers seen annually as 0-5, 6-10, 11-20, or greater than 20. Physicians were queried about estimates of untreated survival time contingent on tumor stage, histologic subtype and grade. Specifically for high and low grade tumors of each histological subtype of epithelial ovarian cancer (serous, mucinous, endometrioid, clear cell and Adenocarcinoma NOS) we asked participants to "Please provide estimates of average survival time in months for patients at each stage at diagnosis and in the absence of treatment: a) invasive ovarian cancer that has metastasized to a distant organ, such as the liver or lung (stage 4), b) invasive ovarian cancer that has spread beyond the ovaries to the abdominal lining, omentum or to lymph nodes (stage 3), c) invasive ovarian cancer that has invaded another organ, but is confined to the pelvis (stage 2), and d) invasive ovarian cancer that is confined to the ovaries (stage 1). A copy of the survey is included as Appendix 1. A total of 80 physicians were sent the survey and 39 responded for an overall response rate of 49%. These responses are summarized in Figure S2.

To calculate stage durations from survival time responses we worked backward from the expected time of death using two models of disease progression – 1) a strict linear progression model whereby the disease progresses strictly from stage 1 through stage 4 prior to death in all patients without skipping any stage and 2) a more relaxed model whereby the disease progresses strictly from stages I through III but allows that a proportion of patients die in stage III without progressing into stage IV. The relaxed model takes into account recent autopsy findings reported by Guth et al. that roughly 43% of patients with untreated ovarian cancer have disease confined to the abdominal-pelvic cavity at the time of death [3]. Because a tumor has presumably been at the stage it is diagnosed in for some time before the clinical (symptomatic) diagnosis and would continue to be at that stage for some time after its clinical diagnosis, an estimate of the portion of the stage that has elapsed at the time of diagnosis is required when calculating pre-clinical disease durations from stage specific expected survivals. We assumed a constant hazard of diagnosis within each stage, and this assumption allowed us to calculate the expected proportion of each stage that has elapsed at the time of the clinical diagnosis based on the probability that the tumor is diagnosed within a stage versus progresses onto the next stage. Grade and histology specific probabilities of progressing through a stage without being diagnosed were estimated from SEER incidence data in combination with the linear progression model. The relaxed disease progression model was used to generate estimates of disease duration and stage lengths presented in Table 2.

e. Age of clinical diagnosis of benign disease: women who are not assigned an age of clinical diagnosis of ovarian cancer may be assigned an age of clinical diagnosis of benign ovarian disease. The incidence of benign disease identifiable by screening is generated by scaling the observed ovarian cancer incidence from the SEER registry by the smoothed ratio of benign to malignant disease reported by Katsube et al. [4]. Women who are assigned neither an age of clinical ovarian cancer diagnosis or age of benign disease diagnosis make up the healthy population.

f. We assume a normal distribution of benign duration with mean of 9 years and standard deviation of 4.5 years. We estimated benign duration from the observed benign cases in years 0-2 of the PLCO screening trial [5]. To calculate prevalence and incidence we used the following recursive relationships: 1) the prevalence of benign disease at any screen is equal to (the prevalence at the prior screen) – (positive rate at the prior screen) + (the annual incidence) and 2) the positive rate at the current screen is the equal to the sensitivity of the test * the prevalence. At the initial screen (year 0), duration = prevalence/incidence.

3. Performance of Screening Modalities

a. A major challenge to empirical modeling of a ovarian cancer screening strategies is the limited amount of data available that can be use to characterize the behavior of screening tests during the relevant period prior to clinical detection of disease. There are few cohorts that are large enough, have been adequately characterized and followed and that have systematically collected pre-diagnostic blood samples and/or screening results from a sufficient number of ovarian cancer cases and controls. We evaluated available data from 3 independent cohorts for incorporation in our analysis.

b. Blood biomarkers: We evaluated two sources of data on the performance of blood biomarkers prior to clinical diagnosis of ovarian cancer. The NCI's Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) which began in 1992 has archived up to six annual blood samples from 155,000 study participants; data on CA125 levels in diagnostic and proximate blood samples from 117 ovarian cancers observed in the PLCO participants were provided to study investigators [5]. The Carotene and Retinol Efficacy Trial (CARET), though smaller than the PLCO, enrolled 18,314 patients at high risk for lung cancer between 1983 and 1994 and followed them for at least six years; there were 34 ovarian and fallopian tube cancers in this cohort [6]. Age distributions for the ovarian cancers in the CARET cohort are as follows: median age 62.5, range: 52 to 79 and SD 6.99 years. We have previously measured CA125 and several other candidate biomarkers in serial pre-clinical samples in cases and controls from this cohort.

Both datasets were analyzed in the same way to estimate the sensitivity of CA125 as a function of time prior to diagnosis. Using a 95% specificity cutoff, the Parametric Empirical Bayes (PEB) algorithm [7] was applied to the CA125 levels of each individual woman to determine if her CA125 was positive or

negative at each observation. The sensitivity of CA125 for 4 time intervals (0-1 year prior to dx, 1-2 years prior to dx, 2-3 years prior to dx and greater than 3 years prior to dx.) was calculated as the percentage of positive tests (as determined by the PEB algorithm) in the given interval.

Although the PLCO dataset has many more observations than the CARET cohort, we were concerned that the CA125 sensitivity function generated using the PLCO data might be biased due to the CA125 screening that was going on in the trial. This screening is expected to bias the sensitivity function towards a shorter overall lead time and increased sensitivity near the time of diagnosis and reduced sensitivity 1 or more years pre-diagnosis Such a trend is observed by comparing PEB sensitivity functions generated using the PLCO and CARET datasets. As shown in Figure S3, the estimated sensitivity in the year immediately before diagnosis is higher in the PLCO data (76%) compared to the CARET data (67%) and is lower at all other time periods in the PLCO data. For this reason, we elected to use more sparse but less biased CARET data to generate the CA125 sensitivity function.

c. Blood based biomarkers are falsely positive in women without a tumor at the time of the blood test at a rate equal to 1- specificity (i.e. the false positive rate, FPR). The FPR was identical for all women and in the absence of a tumor each blood test result is independent. In the presence of the tumor, a semi-deterministic model is used based on the sensitivity profile estimated from the CARET study: when the estimated CA125 sensitivity function threshold for an undetected cancer was less than or equal to the FPR, the chance of a positive CA125 test was identical to that of the non-tumor component (the stochastic FPR) which is the non-deterministic part of this model; in the period when estimated sensitivity function of CA125 was greater than the underlying FPR, a single uniformly generated "detectability" parameter was assigned to the woman and if the estimated sensitivity at the time of the test exceeded the woman's detectability parameter, the test was considered positive, which is the deterministic portion of the model. The semi-deterministic model has the advantage that the proportion of women for whom a marker is positive at some time prior to diagnosis is the same as the sensitivity estimated from the CARET data, and also avoids frequency bias, whereas more frequent screening increases the apparent sensitivity of the screening marker.

4. 2^{nd} line Screens: The 2^{nd} line screen currently implemented uses a fixed sensitivity and specificity; for the base-case analysis TVS, sensitivity and specificity are set to 63% and 97%, and for HI they are set to 90% and 97%. Though sensitivity and specificity of imaging and blood based markers are likely related to many physiological factors such as tumor volume, stage, grade, histology, etc. empirical estimates for these factors and their relationship to current or novel markers cannot be estimated in humans and such relationships are thus not included in this model. An important distinction of this model is that it is aimed primarily at modeling disease related outcomes and not the underlying disease biology. Although both approaches have strengths and limitations we chose to build our model on disease outcomes as at the present time relevant input parameters are more readily observable and empirically driven than for a biologically based model.

5. Screening Strategy
   a. Women are screened annually between ages 45 (starting in 2010) and 85. Other screening schedules are supported by the model, including semi-annual and biannual. Screening continues until diagnosis of ovarian cancer or benign ovarian disease, a false positive screen resulting in removal of ovaries and fallopian tubes or death from other cause prior to age 85. Women not surviving to age 45 are not included in the cohort.
   b. A screening result is considered positive if both the 1^{st} line and 2^{nd} line screen tests are positive.

6. We make the assumption that a positive screen result (positive 1^{st} AND 2^{nd} line screen) leads to surgery in all cases (either salpingo-oophorectomy or laparoscopy/laparatomy). In women destined to have ovarian cancer, BSO surgery prior to the inception of Stage 1 ovarian cancer is assumed to be curative by preventing the disease.

7. Survival

a. For women diagnosed with cancer, a survival time is drawn from distributions conditional on age, stage, grade and histology at diagnosis as estimated from SEER (see Figure S4)
b. For a woman with cancer detected by screening at an earlier stage than stage at clinical diagnosis, a new survival time is drawn from distributions conditioned on age, stage, grade and histology at screen diagnosis. Because the survival distributions are estimated from clinically diagnosed rather than screen-detected cases, the new survival time is adjusted by the difference between her age at clinical diagnosis and her age at screen diagnosis. If the new survival draw does not reach the original survival draw for clinical diagnosis, the original survival time is used (see Figure S4).

8. Cost
   a. Cost of the 1st line screen (CA125 blood test) is incurred at each screen. Screening continues until diagnosis of ovarian cancer or benign ovarian disease, a false positive screen resulting in removal of ovaries and fallopian tubes or death from other cause prior to the end of the screening period. Participant drop out is not modeled, even though some reports have indicated that screening participants are likely to drop out of screening protocols after experiencing a false positive test [5].
   b. The 2nd line screen is modeled as a reflex test that is conducted any time the 1st line screen is positive; cost of the 2nd line screen is incurred only when the 1st line screen is positive.
   c. The cost of a single blood test is $31, as reported by Havrilesky et al. [8] and inflated to 2010 dollars using the Medical care section of the Consumer Price Index . Similarly, the cost of a single ultrasound test is $111 and laparoscopy cost is $4,206.
   d. Costs of treatment are incurred anytime both 1st and 2nd line screens are positive, and are conditional on the woman's disease state at the time of the "positive screen" (i.e. both first and 2nd line tests are positive).
   e. For women who lack a malignant ovarian tumor at the time of a positive screen, treatment is assumed to include only laparoscopic BSO. In this case, the cost of laparoscopy is applied, regardless of whether not the woman would have eventually been diagnosed with a malignant ovarian tumor in the absence of screening.
   f. Treatment for clinically diagnosed benign ovarian conditions includes only laparoscopic BSO. For women diagnosed with benign ovarian disease in the absence of screening, the cost of laparoscopy may be shifted to an earlier date in the presence of screening but is not increased or avoided relative to the no-screening scenario.
   g. For women with cancer, treatment costs (both in the presence and absence of screening) consist of three parts; cost of initial twelve months, cost of continuing care (annual), cost of last year of life. Fractional costs can be incurred if survival is less than one year. As an example, a patient surviving 4 years and 4 months from the time of diagnosis would receive 1 year of diagnosis costs, 2 years 4 months of costs of continuing care costs and 1 year of last year of life costs. All costs are spread uniformly over the associated survival period and discounted on a continuous scale.

      These costs are estimated by inflating the 2004 Medicare costs reported in Yabroff et al. to 2010 dollars using the medical care section of the Consumer Price Index . We only include medical care costs in our cost estimates; our calculations do not include other types of economic impact such as lost wages or reduced productivity.

9. Discounting
   a. All costs are reported in 2010 dollars. Dollar costs and years of life saved are uniformly and continuously discounted from the year of enrollment (2010) at a rate of 3% per year using the following formula where the "cost" variable may be denominated in years of life or 2010 dollars:
   b. a = Year of enrollment; b = Year cost is incurred

      $$\text{Discounted Cost} = (\text{Cost}/0.03)*\exp(-a*0.03) - \exp(-b*0.03)$$

10. Technical Details
    a. Pseudo random number streams which produce a unique and repeatable sequence of values are used to generate values for each attribute that varies over the lifetime of a woman (marker levels, whether or not an Ultrasound will be positive on a given day, etc). Each stream can be uniquely determined by its

"seed" and each property which varies over time within women are assigned a unique seed for that woman.

b. Weighting is used to reduce model computation time in order to direct computational resources more efficiently. Because ovarian cancer cases make up a tiny fraction of the entire population, the greatest limitation to the precision of the outcomes is the number of cases. By only generating life histories and associated cost estimates for a fraction of the healthy women included in the cohort, we are able to increase the precision of the model within a fixed computation time/cost. Healthy women are weighted by 30 (meaning that we only generate $1/30^{th}$ of the healthy women included in the cohort) and women with benign disease are weighted by 3. Women with cancer are weighted by 1, that is, unweighted.

11. Model Validation

a. To validate the computation of treatment costs for cancer we compared the average cost of treatment as estimated by the model with the expected cost given average number of treatment years. Without screening, average cost of treatment for women who died of their cancer after at least two years of cancer treatment, by stage of disease (local=Stage 1, regional = Stage II, distant = Stage III and IV), as estimated by the model are $85,320, $118,649 and $151,777, respectively. The corresponding average number of cancer treatment years are 6.48, 6.56 and 4.55. Applying the model treatment costs described in Table 2 to the average number of cancer treatment years from the model results in average treatment costs of $85,318, $118,646 and $151,775 for local, regional and distant disease, respectively. Similar calculations were performed for women who died of other causes or received cancer treatment for less than two years (data not shown).

b. To check model computation of first and second line screen performance, we performed several model runs where the parameters of interest were set to boundary conditions and verified that the model outputs behaved as expected (see Table S3 for the complete output). Fixing the sensitivity of the first line screen to 0% and the specificity to 95% at all times prior to diagnosis resulted in a 5% mortality reduction. Fixing the sensitivity of the second line screen at 0% and the specificity at 97% resulted in a 3% mortality reduction.

References

1. U.S. Department of Health and Human Services. PHS Centers for Disease Control. NCHS 1991. VitaI Statistics of the U.S. 2988 Vol. II Mortality, Part A, Table 14. Sect. 1, General Mortality, 7.
2. Surveillance, Epidemiology, and End Results (SEER) Program (http://seer.cancer.gov/) SEER*Stat Database: Incidence - SEER 9 Regs Limited-Use, Nov 2006 Sub (1973-2004) - Linked To County Attributes - Total U.S., 1969-2006 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2007, based on the November 2006 submission.
3. Guth U, Huang DJ, Bauer G, Stieger M, Wight E, Singer G. Metastatic patterns at autopsy in patients with ovarian carcinoma. Cancer. 2007 Sep 15;110(6):1272-80.
4. Katsube Y, Berg JW, Silverberg SG. Epidemiologic pathology of ovarian tumors: a histopathologic review of primary ovarian neoplasms diagnosed in the Denver Standard Metropolitan Statistical Area, 1 July-31 December 1969 and 1 July-31 December 1979. Int J Gynecol Pathol. 1982;1(1):3-16.
5. Partridge E, Kreimer AR, Greenlee RT, Williams C, Xu JL, Church TR, et al. Results from four rounds of ovarian cancer screening in a randomized trial. Obstet Gynecol. 2009 Apr;113(4):775-82. PMC2728067
6. Anderson GL, McIntosh M, Wu L, Barnett M, Goodman G, Thorpe JD, et al. Assessing lead time of selected ovarian cancer biomarkers: a nested case-control study. J Natl Cancer Inst. 2010 Jan 6;102(1):26-38. 2802285
7. McIntosh MW, Urban N, Karlan B. Generating longitudinal screening algorithms using novel biomarkers for disease. Cancer Epidemiol Biomarkers Prev. 2002 Feb;11(2):159-66.
8. Havrilesky LJ, Sanders GD, Kulasingam S, Myers ER. Reducing ovarian cancer mortality through screening: Is it possible, and can we afford it? Gynecol Oncol. 2008 Nov;111(2):179-87.

9. Appendix: 2003 Statistics on All-listed Procedures in U.S. Hospitals. Available from: http://stats.bls.gov/cpi/home.htm#data.