Research Article

Cancer Prevention Research

Impact of Screening Test Performance and Cost on Mortality Reduction and Cost-effectiveness of Multimodal Ovarian Cancer Screening

Charles W. Drescher[1], Sarah Hawley[2], Jason D. Thorpe[1], Simone Martícke[2], Martin McIntosh[1], Sanjiv S. Gambhir[3], and Nicole Urban[1]

Abstract

Ongoing ovarian cancer screening trials are investigating the efficacy of a two-step screening strategy using currently available blood and imaging tests [CA125 and transvaginal sonography (TVS)]. Concurrently, efforts to develop new biomarkers and imaging tests seek to improve screening performance beyond its current limits. This study estimates the mortality reduction, years of life saved, and cost-effectiveness achievable by annual multimodal screening using increasing CA125 to select women for TVS, and predicts improvements achievable by replacing currently available screening tests with hypothetical counterparts with better performance characteristics. An existing stochastic microsimulation model is refined and used to screen a virtual cohort of 1 million women from ages 45 to 85 years. Each woman is assigned a detailed disease course and screening results timeline. The preclinical behavior of CA125 and TVS is simulated using empirical data derived from clinical trials. Simulations in which the disease incidence and performance characteristics of the screening tests are independently varied are conducted to evaluate the impact of these factors on overall screening performance and costs. Our results show that when applied to women at average risk, annual screening using increasing CA125 to select women for TVS achieves modest mortality reduction (~13%) and meets currently accepted cost-effectiveness guidelines. Screening outcomes are relatively insensitive to second-line test performance and costs. Identification of a first-line test that does substantially better than CA125 and has similar costs is required for screening to reduce ovarian mortality by at least 25% and be reasonably cost-effective. *Cancer Prev Res; 5(8); 1015–24. ©2012 AACR.*

Introduction

The impact of epithelial ovarian cancer (EOC) screening using the serum tumor marker CA125 and transvaginal sonography (TVS) is being evaluated in 2 large efficacy trials. The Prostate Lung Colon and Ovary (PLCO) trial in the United States failed to show a reduction in EOC-specific mortality with annual screening using CA125 interpreted using a single threshold rule and TVS concurrently (1). The United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) compares 2 screening approaches to no screening (2). In its single-modality imaging arm, TVS alone is used annually; in its multimodal arm, increasing CA125 is used as a first-line screen to select women for a second-line screen with TVS. The multimodal arm uses the longitudinal risk of ovarian cancer algorithm (ROCA; ref. 3), relying on callbacks for repeat CA125 measurement to confirm exponential rise. Women with both rising CA125 and abnormal imaging results are referred for surgical consult. Early results of the trial suggest that the CA125-driven multimodal strategy outperforms the single-modality strategy using TVS. The sensitivity, specificity, and positive predictive value for all primary invasive EOCs identified at the prevalence screen in the UKCTOCS were 89.5%, 99.8%, and 35.1%, respectively, for the multimodal strategy, and 75.0%, 98.2%, and 2.8%, respectively, for TVS alone (2). Promising results using this same multimodal strategy were also recently report by Lu and colleagues (4). Although promising, it is not yet known whether the sequential multimodal screening strategy will reduce mortality cost-effectively.

Efforts are underway to identify biomarkers and imaging strategies that perform better than currently available tests. Clinical studies have evaluated the potential of several serum marker candidates for EOC screening (5, 6), and although some have showed a detection lead time of about a year in retrospective validation studies (7, 8), none seem to perform better than CA125 either alone or as a panel. Motivated by advances in molecular biology,

Authors' Affiliations: [1]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington; [2]Canary Foundation, Palo Alto; and [3]Stanford University, Stanford, California

Note: Supplementary data for this article are available at Cancer Prevention Research Online (http://cancerprevres.aacrjournals.org/).

Corresponding Author: Charles W. Drescher, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98104. Phone: 206-965-1750; Fax: 206-965-1735; E-mail: cdresche@fhcrc.org

doi: 10.1158/1940-6207.CAPR-11-0468

nanotechnology, and imaging technology, molecular imaging strategies are rapidly approaching the resolution necessary for EOC early detection (9). Molecularly targeted microbubble ultrasound contrast agents are a particularly appealing approach for EOC screening because they can use conventional ultrasound technology, a widely available and relatively inexpensive ovarian imaging method. Small animal imaging experiments show that microbubbles targeted to VEGF receptor (VEGFR2) expressed by activated endothelial cell substantially improve ultrasound signal intensity and resolution (9–11). VEGFR2-targeted microbubbles are currently undergoing pilot testing in patients.

Knowledge about the relative effectiveness and cost-effectiveness of both conventional and new screening tools will help guide strategies for technology development and deployment toward large population EOC screening interventions, which are costly and resource intensive. Simulation modeling can be an effective and efficient method for evaluating alternative screening strategies, allowing comparison of the performance of many screening strategies under various assumptions about disease behavior and screening test performance. Modeling has been used to predict mortality reduction and cost-effectiveness of multimodal EOC screening using CA125 and TVS as first- and second-line tests, respectively (12, 13).

It is unknown what performance characteristics and cost parameters will be required of new EOC screening tests to have a significant impact on EOC mortality and be cost-effective. Here, we refine and extend a previously developed (14) and validated (13) EOC screening microsimulation model and use the model to address these important questions. Our results confirm earlier observations that currently available screening tests are likely to have only a modest impact on EOC mortality but are reasonably cost-effective. We further show that the performance of a multimodal EOC screening strategy is largely dependent on the characteristics of the first-line test. Identification of a first-line screening test that performs substantially better than CA125 and costs no more than $95 will be required to have a major impact on EOC mortality and be reasonably cost-effective when conducted annually in average-risk postmenopausal women.

Materials and Methods

Overview

A stochastic microsimulation is used to estimate the mortality reduction, years of life saved (YLS), and cost-effectiveness of EOC screening protocols in a hypothetical cohort of 1 million women beginning at age 45 and continuing through age 85. Four distinct components comprise the model to represent (i) natural history for women with and without EOC (Natural history component), (ii) EOC screening protocols and test results (Screening component), (iii) EOC survival adjusted for age, stage, and histology at diagnosis (Survival component), and (iv) costs associated with EOC screening, diagnosis, and treatment including costs incurred as a result of false-positive screens (Cost component). The model generates disease-related outcomes and costs for women in the absence of screening, then superimposes a screening strategy on the cohort and calculates the impact of screening on survival and costs. Mortality reduction is measured as the decrease in deaths due to EOC divided by the EOC deaths that would have occurred in the absence of screening among women with disease present during the screening period. YLS are reported as difference in age of death with and without screening. All costs including screening, diagnostic, and treatment costs are reported in 2010 U.S. dollars, and all future costs and benefits are discounted back to 2010 using a 3% rate of return. The model calculates YLS and costs for each woman and reports both cumulatively for the entire screening cohort (for additional information, see Supplementary Methods Section 1).

Empirical data obtained or generated from experimental analysis, public use files, or published literature provide input parameters for the model (Table 1). Where empirical data are sparse or unavailable (e.g., novel and hypothetical screening modalities), we apply our best estimates and test their robustness using sensitivity analyses. The model is stochastic in that it includes a component of randomness to represent "luck" for an individual woman. Below, we highlight key features and input data for each component comprising the model.

Natural history component

The natural history component generates a cohort of women with ages at death and incidence of EOC, using competing risks of developing EOC or dying cancer free that are derived from Surveillance Epidemiology and End Results (SEER; ref. 15) and the U.S. Vital Statistics Report (16). The generated cohort is divided into 4 groups: women with symptomatic EOC (cases), women with nonmalignant ovarian tumors (benigns), women without ovarian disease (healthies), and women who develop an occult EOC but die from competing causes before clinical diagnosis (latents). For EOC cases and latents, tumor characteristics including age at clinical diagnosis, stage, histology, and grade are derived to be consistent with U.S. nationwide incidence rates reported by SEER (15). The date of cancer inception, defined as the earliest time when the tumor is pathologically malignant, is calculated backward from the date of clinical diagnosis using estimates of disease duration and stage lengths obtained by surveying gynecologic oncologists (Table 2).

Benign disease distributions are derived from Katsube and colleagues (17), who report the incidence for primary ovarian tumors (all histologies) identified among women in the Denver Standard Metropolitan Statistical Area over a 10-year period. Similar to EOC cases, age at benign disease inception is back calculated from age at clinical diagnosis assuming mean disease duration of 9 years, an estimate calculated from prevalence and incidence rates of benign ovarian tumors identified among participants in the PLCO Trial (see Supplementary Methods Section 2.f).

Table 1. Model inputs

Input parameter	Disease group	Baseline assumption	Source
Natural History component			
Age at death from competing cause	All	N/A	Vital statistics of the United States (16)
Incidence and age at clinical diagnosis for cases	Cases, Latent	N/A	SEER (15)
Incidence and age at clinical diagnosis for benign disease	Benign	N/A	Katsube and colleagues (17)
Tumor characteristics (stage, histology, and grade at clinical diagnosis)	Cases, Latent	N/A	SEER (15)
Malignant disease duration and stage lengths	Cases, Latent	See Table 2	
Benign disease duration	Benign	9 years	PLCO (26)
Screening component			
Screening frequency	All	Annual (age 45–85 years)	N/A
CA125 sensitivity	Cases, Benign	See Fig. 2	CARET (7)
CA125 specificity	Healthy	95%	Defined by the screening algorithm
Hypothetical marker sensitivity	Cases, Benign	2× sensitivity of CA125	
Hypothetical marker specificity	Healthy	95%	
TVS sensitivity	Cases	63%	PLCO (26)
TVS specificity	Healthy	97%	PLCO (26)
Hypothetical imaging sensitivity	Cases	90%	
Hypothetical imaging specificity	Healthy	97%	
Survival component			
EOC survival contingent on age and tumor stage, histology, and grade at diagnosis	Cases	See Supplementary Figure S4	SEER (19)
Cost component[a]			
CA125 test cost	All	$31	Havrilesky and colleagues (12)
Hypothetical marker test cost	All	$210	
TVS test cost	All	$111	Havrilesky and colleagues (12)
Hypothetical imaging test cost	All	$750	
Laparoscopy with BSO	Benign, Healthy	$4,206	Havrilesky and colleagues (12)
EOC treatment costs	Cases		Yabroff and colleagues (21)
Initial year	Stage I	$36,672	
	Stage II	$50,719	
	Stages III/IV	$70,452	
Continuing care	All stages	$4,712	
Last year of life[b]	Stage I	$27,523	
	Stage II	$46,438	
	Stages III/IV	$69,314	

[a]All costs were adjusted to 2010 U.S. dollars.

[b]The treatment cost differential for the last year of life between early and late stage diagnoses arises by the way Yabroff and colleagues allocated treatment costs for patients surviving less than 24 months past diagnosis. Costs for the last 12 months of this period were allocated as last year of life costs, and the remainder considered initial year treatment costs. Our model treats cost allocation for such patients in a similar fashion. Allocation of death-related costs may increase initial year treatment costs for women who die within 12 months of diagnosis.

Table 2. Mean disease duration and stage lengths in years for EOC cases by histology and grade

Low-grade					
Stage length	**Serous**	**Mucinous**	**Endometrioid**	**Adenocarcinoma NOS**	
Stage 1	1.8	2.2	1.7	2.3	
Stage 2	0.9	0.2	0.9	0.2	
Stage 3	1.4	1.4	1.4	1.5	
Stage 4	0.5	0.5	0.5	0.5	
Total disease duration[a]	4.6	4.3	4.5	4.5	
High-grade					
Stage length	**Serous**	**Mucinous**	**Endometrioid**	**Adenocarcinoma NOS**	**Clear cell**
Stage 1	1.0	0.6	1.0	1.5	0.9
Stage 2	0.4	0.6	0.6	0.2	0.3
Stage 3	0.8	0.8	0.8	0.9	0.8
Stage 4	0.2	0.3	0.3	0.2	0.2
Total disease duration[a]	2.5	2.3	2.6	2.8	2.2

NOTE: Calculated using physicians' estimates of EOC progression rates in the absence of treatment (detailed in Supplementary Methods Section 2.d.)

[a]Total disease duration was modeled using the lognormal distribution with 0.5 year standard deviation, and then divided into stages using the Dirichlet distribution to yield stages with the above average durations.

Screening component

The screening component generates for each woman a complete timeline of screening test results, superimposes an EOC screening protocol onto the cohort, and calculates shifts in disease detection from late- to early-stage attributable to screening. Two types of screening tests are modeled: serum biomarkers and imaging. As a base case scenario, we employ a 2-step screening strategy using CA125 to select women for TVS (CA125 + TVS). Women are screened annually from age 45 through 85 years. CA125 test results are interpreted using the parametric empirical Bayes (PEB) rule, a longitudinal algorithm previously described (18). The PEB uses PEB statistical theory to generate person-specific positivity thresholds that depend on the screening history of each individual while holding the false-positive rate constant across all of the screened population. Any predefined specificity level can be used. We chose the 95% specificity level for our analysis, meaning that at each screen 5% of healthy women are expected to receive a positive test result.

For EOC cases and latents, CA125 test results are assigned based on a sensitivity function that provides an estimate for the probability of a positive PEB CA125 test result at selected time intervals before clinical diagnosis (Fig. 1). The sensitivity function is derived from analysis of CA125 levels in preclinical samples collected from ovarian cases identified among Carotene and Retinol Efficacy Trial (CARET) trial participants (see Supplementary Methods Section 3.b). Sensitivity increases with proximity to diagnosis; the baseline 5% positive test threshold (corresponding to 95% specificity) is applied at times remote from diagnosis when the function decreases below this threshold. A sensitivity function for benign ovarian tumors is derived by scaling the EOC case function by a factor of 0.15, an adjustment that closely approximates the incidence of positive PEB CA125 tests attributable to benign tumors identified among 328 participants in a local EOC screening protocol (data not shown). The imaging component of the model assumes that TVS is equally sensitive throughout the disease duration once CA125 values have elevated above the positivity

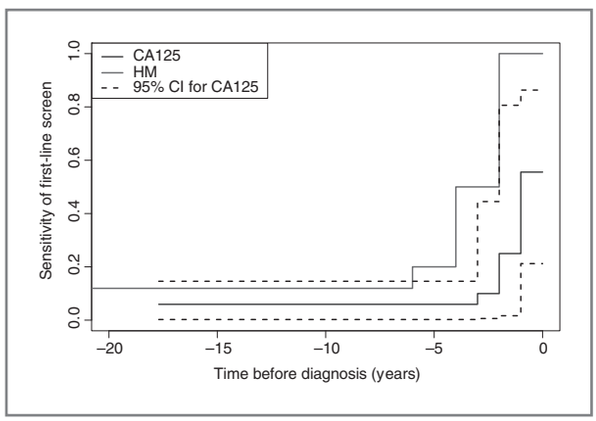

Figure 1. Ovarian cancer case sensitivity function for PEB CA125 and HM by time before clinical diagnosis. The CA125 sensitivity function is derived by applying the PEB algorithm to serial CA125 levels measured in preclinical samples obtained from Carotene and Retinol Efficacy Trial Participants (see Supplementary Methods Sec 3). The baseline 5% positive test result corresponding to 95% specificity is applied at times remote from diagnosis when the function decreases below this threshold. The sensitivity function for HM was derived by doubling both the sensitivity and lead-time of CA125. The 95% confidence interval for each piece of the CA125 sensitivity function is estimated using the exact binomial method.

threshold. Because TVS is used as a second-line screen, no assumptions on the performance of TVS are made when CA125 is not elevated. We chose 63% and 97% as baseline input parameters for TVS sensitivity and specificity, respectively, based on data from the PLCO trial as this represents the current best estimate for the performance of community-based sonography; we also conduct sensitivity analysis across a range of assumptions regarding the sensitivity of TVS (Supplementary Table S7). We assume that all women with both elevated markers and abnormal imaging undergo immediate diagnostic surgery. Requiring positivity for both tests yields a probability of surgical referral among healthy women of 1.5/1,000 at each screen.

A laparoscopic bilateral salpingo-oophorectomy (BSO) is conducted in women referred for surgery with benign ovarian tumors or normal ovaries at the time of the screen (false-positive screen); risk of developing EOC following BSO is assumed to be zero. EOC cases are assumed to undergo appropriate EOC surgery followed by other clinically indicated therapy. The date at screen diagnosis is defined as time of the positive screen after disease inception. Whenever the date of screen detection precedes the date at clinical diagnosis, the tumor stage at screen diagnosis is calculated by applying tumor-specific estimates of stage durations (Table 2) to the date of tumor inception. A stage shift is assumed to occur whenever a tumor destined to be diagnosed clinically in late stage (III or IV) is detected in early stage (I or II) by screening. For each screen-detected case, we calculate the lead time as the time interval between the first true positive screen and the date when clinical diagnosis would have occurred in the absence of screening. The average lead time among screen-detected cancers is reported as an output from the model.

Performance characteristics of hypothetical screening modalities. To estimate the impact of potential improvements in EOC screening tools, we simulated a hypothetical biomarker (HM) with 2-fold greater sensitivity and lead time than that of CA125 (Fig. 1). These performance characteristics are arbitrary and intended to represent an optimistic appraisal of what is potentially achievable over the next decade with new biomarker discovery efforts. A sensitivity function of the HM for benign disease is derived by scaling the HM sensitivity function for EOC cases by 0.15. We also model the performance of a hypothetical imaging test (HI) that achieves a 50% improvement in sensitivity relative to TVS (from roughly 60%–90%) while maintaining high specificity (97%). The 90% sensitivity for HI is assigned uniformly at all times following disease inception, representing a significant improvement in imaging resolution especially when used as a confirmatory test following a HM with 2× longer lead time and sensitivity compared with CA125.

Survival component

The survival component generates a date of death for all EOC cases detected clinically or by screening taking into account disease-specific survival and competing cause mortality. EOC survival curves extending out to 15 years postdiagnosis specific to age at diagnosis (in 10 year intervals), stage (I, II, and stages III and IV combined) and grade (low vs. high) were estimated from SEER (19) and are used to generate survival after EOC diagnosis (Supplementary Fig. S4). Age at death is set to the earliest of competing risk and disease-specific mortalities. Cases experiencing a false-positive screening result and BSO before tumor inception are assumed to be protected from an EOC diagnosis. Women alive 15 years after EOC diagnosis are assumed to be cured with age of death determined by competing risks. We include a 0.1% risk of death associated with surgical investigation of a false-positive screening result (20). For established malignancies, surgical deaths are accounted for in the SEER survival curves (19). We assume all women die by age 110. The overall impact of screening on EOC-specific survival is reported as the total number of YLS and mortality reduction (the percentage of women saved from dying of EOC) across the entire cohort.

Cost component

The cost component calculates the cost of the screening program and interventions associated with EOC diagnosis and treatment under both screening and nonscreening scenarios. Three categories of costs are considered: (i) cost associated with the screening tests themselves, (ii) costs associated with surgical evaluation of positive screens, and (iii) cancer treatment costs. Medicare reimbursement rates taken from prior reports and based on Medicare claims data are used for input cost parameters wherever possible (Table 1).

EOC treatment costs are based on Yabroff and colleagues (21) who estimate net EOC treatment costs by stage and phase of disease using 1999–2003 Medicare Part A and B claims data on 1,647 EOC cases identified from SEER and matched controls. Treatment costs are divided into 3 components: (i) costs incurred in the first year of diagnosis (including primary surgery costs), (ii) costs incurred in the final year of life before death, and (iii) continuing care costs for all years between diagnosis and death. The higher initial and terminal treatment costs for cases diagnosed at stage III/IV reflect the additional expenses associated with treating more advanced disease. We assume that a woman surviving 15 years postdiagnosis is cured and incurs no additional costs. All costs are reported in 2010 dollars and future costs and YLS are discounted to 2010. The cost-effectiveness (cost per YLS) of the screening program is calculated as the net cost of screening (screening costs − saving in treatment costs due to early disease detection) divided by the YLS attributable to screening.

The characteristics and costs of hypothetical screening tests capable of achieving the defined performance parameters are largely unknown. For our base case analysis, we assume the HM and HI cost roughly 7× their current counterparts (CA125 and TVS) yielding a cost of $210 for HM and $750 for HI. The estimated cost for HM is roughly equivalent to current charges for HE4, a new FDA-approved EOC serum marker. The cost for HI is consistent with authors' estimated lowest possible charge for targeted

microbubble contrast agents when used in conjunction with a TVS screen. We also report the overall cost-effectiveness of implementing these tests across a range of assumptions of screening test costs.

Results

Mortality reduction

Mortality reduction was estimated for 4 different annual multimodality screening strategies under base case assumptions (Table 3). Screening using increasing CA125 followed by TVS achieves a mortality reduction of 13%. Substituting HI for TVS as the second-line screen improves mortality reduction only modestly to 15%. Greater mortality reduction (25%) is achieved by using HM in place of CA125 as the first-line screen before TVS. Separately implementing the sensitivity and the lead-time improvement associated with HM leads to a mortality reduction of 19% and 17%, respectively (data not shown). Screening using both HM and HI achieves the greatest mortality reduction at 30%.

Because we assume that a positive screen necessitates BSO even in the absence of EOC, surgical evaluation of false-positive screens can potentially prevent EOC in women destined to develop EOC later. Consequently, we separately evaluated the proportion of the overall mortality reduction attributable to early detection of established disease versus disease prevention from false-positive surgeries. The proportion of overall mortality reduction attributable to early detection of established disease increases from roughly 70% for CA125 + TVS to 86% for the best performing screening strategy of HI + HM. Absolute mortality reduction attributable to disease prevention is 4% across all 4 screening scenarios as the overall specificity (and hence false-positive rate) of the screening programs are identical.

Cost-effectiveness

Annual multimodal screening using CA125 + TVS yields a cost of $88,993 per YLS (Table 3). Implementing HI in place of TVS at a cost of $750 per test saves more lives but is less efficient costing $124,376 per YLS. Screening using HM at a cost of $210 has a major impact on mortality reduction but substantially increases cost regardless of the second line test. Overall cost per YLS of the HM + TVS and HM + HI strategy is $205,248 and $191,441 respectively. HI is more cost-effective than TVS because cost savings from the increase in YLS exceed the additional costs associated with the test.

Our assumptions about the costs of the hypothetical screening tests affect the overall cost-effectiveness of the screening strategies (Fig. 2, Supplementary Fig. S4). The

Table 3. Mortality reduction, years of life saved (YLS) and cost per YLS of EOC screening strategies

First-line screen	**None**	**CA125**	**CA125**	**HM**	**HM**
Second-line screen	**None**	**TVS**	**HI**	**TVS**	**HI**
Effectiveness					
Stage at Dx:					
I	18%	34%	39%	57%	67%
II	11%	13%	13%	14%	12%
III	32%	30%	28%	21%	16%
IV	39%	23%	19%	8%	4%
Stage shifts (late to early)[a]	N/A	17%	22%	40%	48%
Total mortality reduction	**N/A**	**13%**	**15%**	**25%**	**30%**
Due to early detection of malignancy in cases	N/A	9%	11%	21%	26%
Due to BSO in cases before the onset of malignancy	N/A	4%	4%	4%	4%
Lead time (per screen detected case)	N/A	0.97	0.97	1.48	1.60
YLS (per screen detected case)	N/A	1.68	1.61	2.09	2.21
Cost					
Cost per YLS	**N/A**	**$88,993**	**$124,376**	**$205,248**	**$191,441**
Total cost (millions)	$865	$1,741	$2,397	$5,401	$6,068
Screening test cost (total in millions)	$0	$742	$1,396	$4,373	$5,039
Laparoscopy cost (millions):					
False-positive: healthy women	$0	$128	$128	$128	$128
False-positive: benign disease	$81	$84	$83	$86	$83
Cancer treatment cost (millions)	$784	$787	$791	$814	$819

[a]Minor inconsistencies between the proportion of women experiencing a screening related stage shift and changes in stage distribution relative to the no screen control group are attributable to ovarian cancer prevention in a small number of screened women who undergo BSO before tumor development as a consequence of a false-positive screen. This prevention reduces the number of ovarian cancer cases overall and the denominator used in the stage shift and distribution calculations.

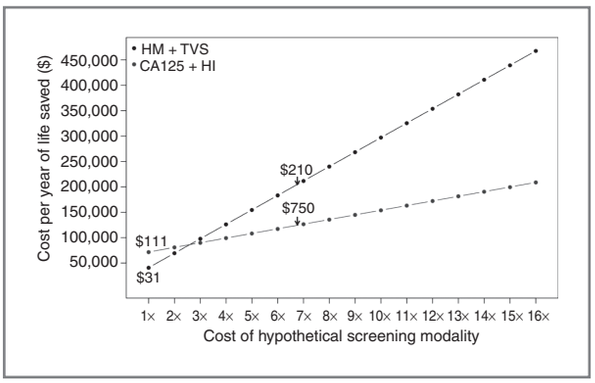

Figure 2. Cost-effectiveness of screening using hypothetical screening tests by screening test cost. Hypothetical screening test costs are presented as fold cost relative to the cost of CA125 ($31) and TVS ($111). $210 and $750 correspond to the base case assumptions for the cost associated with HM and HI, respectively.

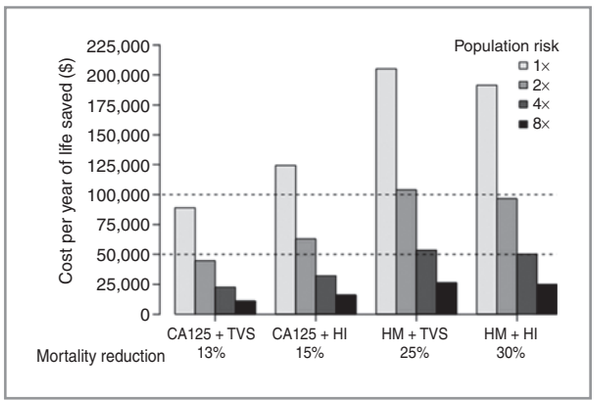

Figure 3. Cost-effectiveness of ovarian cancer screening strategies in populations at increased risk. We evaluated the cost-effectiveness of all 4 bimodal screening strategies when applied to populations with increased ovarian cancer incidence rates of 2X, 4X, and 8X relative to the general population.

overall cost-effectiveness of replacing TVS by HI was $71,772 for a HI with a cost equivalent to that of TVS ($111), increasing by $9,128 YLS for each fold increase in test cost. Replacing CA125 by HM results in cost per YLS of $40,926 for a HM with a cost equivalent to that of CA125 ($31), increasing by $28,458 for every fold increase in the test cost. The cost per YLS of HM+TVS remains below the generally acceptable threshold of $100,000 provided the cost of HM is $95 or less. The cost per YLS estimated for CA125+HI strategy remains below the $100,000 YLS threshold provided the cost of HI is no greater $454 per test.

Cost-effectiveness is improved by screening populations at increased risk. We evaluated the cost-effectiveness of annual screening using each strategy when applied to populations with EOC incidence rates of 2×, 4×, and 8× relative to the general population (Fig. 3, Supplementary Table S2). As expected, the cost-effectiveness of all the screening programs improves in proportion to the increase in disease incidence because, as more women are screen-detected in elevated risk populations (increased pretest likelihood), the increase in number of YLS exceeds the increase in screening costs. The mortality reduction for a given screening program is unaffected by disease incidence because the performance characteristics of the screening tests do not change.

Screening frequency

Screening frequency affects both mortality reduction and cost-effectiveness. We modeled the interaction between screening frequency and costs for each of the 4 screening strategies when applied to an average risk population and compared the results across strategies (Fig. 4, Supplementary Table S3). As expected, mortality reduction and cost per YLS both increase for all screening strategies as screening intervals are shortened. CA125-based screening strategies do not achieve a mortality reduction of 30% even when conducted semiannually. Screening using HM + HI at least annually or HM + TVS semiannually can achieve this target. However, HM–based screening strategies are relatively expensive due to cost of the HM which is applied at every screen.

Model variability and reliability

We evaluated variability in model output by conducting 100 model runs of the CA125 + TVS screening strategy using independent random number streams while holding all model parameters constant. In this analysis, the average (min–max) percent mortality reduction, YLS per screen detected case, and cost per YLS was 12.5% (11.9%–13.3%), 0.64 (0.58–0.70), and $87,664 ($80,072–$96,417), respectively (Supplementary Table S4). This variability is attributable to the stochastic (i.e., random) generation of women's natural history using empirically based random distributions. When different screening strategies are compared, the stochastic element of the model is held constant, so that each strategy is evaluated using an identical cohort of women to make a level comparison.

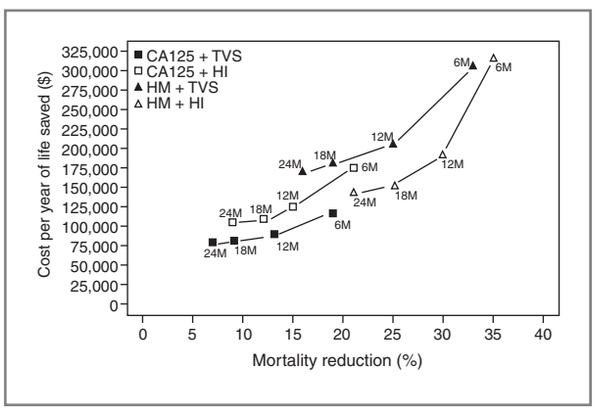

Figure 4. Impact of screening frequency on mortality reduction and cost-effectiveness of ovarian cancer screening strategies. Mortality reduction and cost-effectiveness for screening intervals of 6, 18, and 24 months are shown in addition to the base case assumption of 12 months, for each of the four screening scenarios. All other parameters were held constant.

We also validated the model for consistency. External consistency validation was conducted using boundary conditions of the various input parameters for which model outcomes could be directly calculated external to the model. Internal validations were conducted by verifying that relationships between model outputs (such as "cumulative treatment costs = cumulative survival × survival-dependent treatment costs") met expected relations (Supplementary Table S5 and Supplementary Methods Section 11).

Sensitivity analyses

Sensitivity analysis was conducted to evaluate robustness of our results to halving and doubling input parameters used in the base case (Table 4), particularly for those with limited available empirical data such as disease progression which is not directly observable in humans so that our estimates of EOC stage lengths and disease duration cannot be directly verified.

Doubling our assumptions about the duration of stage I disease does not change estimated mortality reduction and has only a minor impact on cost-effectiveness for screening using CA125 + TVS. This is attributable to the fact that the empirically derived CA125 sensitivity function is conditional on time before clinical diagnosis and not on imputed stage at the time of the screening test. Because CA125 levels begin to increase appreciably only between 1 and 2 years before clinical diagnosis (Fig. 2), they are relatively insensitive to the duration of early-stage disease. Stage I disease duration has a more significant impact on HM–driven screening strategies because of the longer lead-time attributable to HM. For example, for the HM + TVS screening strategy, doubling stage I disease duration increases mortality reduction from 25% to 28% and improves cost per YLS from $205,248 to $189,532.

Although empirically derived our base case assumptions about the performance of CA125 before clinical diagnosis are imprecise and may either under- or overestimate the true sensitivity of CA125. Mortality reduction of 8% and 20% is achieved when the lower and upper bounds of the 95% confidence intervals for the CA125 sensitivity function are used as input parameters for CA125 performance, respectively (see Table S1 Supplemental Materials).

Changes in our assumptions about the performance of TVS considerably impact both mortality reduction and cost-effectiveness. Assigning TVS near perfect sensitivity of 99.9% increases mortality reduction to 16% and reduces cost per YLS to $67,605. Importantly, the 16% mortality reduction associated with 99.9% sensitivity analysis for TVS also defines the upper bounds of what is potentially achievable for a HI when used as follow-up to an elevated CA125. Lowering the specificity of TVS to 94% (doubling the false-positive rate to 6%) increases mortality reduction from 13% to 15% and improves cost-effectiveness because it leads to more diagnostic surgeries in women destined to develop cancer later on.

Discussion

We used an updated microsimulation model to estimate the potential benefits of EOC screening. Simulation of annual screening using rising CA125 to select women for TVS predicts mortality reduction of approximately 13% at a cost of $89,000 per YLS. Semiannual screening increases the mortality reduction to roughly 20% but is less cost-effective at $117,350 per YLS. The mortality reduction we identified is consistent with Havrilesky and colleagues who, using a Markov transition state model, estimated a mortality reduction of 14.7% for an annual sequential screening strategy when applied to a postmenopausal population. Mortality reduction estimated by that model decreased to 10.9% when the authors accounted for variability in EOC aggressiveness. Combined, these data suggest the potential effectiveness of EOC screening using currently available tools is likely to be modest.

The updated model incorporates new functionality and is more empirically driven than that used in earlier work. The natural history component has been expanded to account for heterogeneity in EOC behavior by incorporating representations of disease duration and survival broken down by tumor histology and grade. We now also use empiric data obtained from analysis of CA125 levels in preclinical blood samples rather than a simple exponential model to characterize the trajectory of CA125 before diagnosis. These data show initial elevation in CA125 levels between 1 and 2 years before clinical diagnosis and provide a more realistic approximation of the potential benefit of

Table 4. Impact of varying input parameters on mortality reduction and cost-effectiveness of annual screening using CA125+TVS

Disease duration (Stage I malignant)	**0.5×**	**1×**	**2×**
Mortality reduction	11%	13%	13%
Cost-effectiveness	$98,018	$88,993	$85,556
TVS sensitivity	**31.5%**	**63%**	**99.9%**
Mortality reduction	8%	13%	16%
Cost-effectiveness	$139,779	$88,993	$67,605
False-positive Rate (TVS specificity)	**1.5% (98.5)**	**3% (97%)**	**6% (94%)**
Mortality reduction	11%	13%	15%
Cost-effectiveness	$91,763	$88,993	$83,725

screening using CA125. In our earlier report, we estimated an average lead time for annual CA125-based screening of roughly 28 months which is too optimistic and likely explains the more significant mortality reduction (31%) reported earlier (14).

Our assumptions about EOC disease duration are based on gynecologic and medical oncologists' estimates of disease progression in the absence of therapy. Because they cannot be validated, we evaluated their impact on our results using sensitivity analysis. We focused on stage I disease duration as this is likely to have the greatest impact on screening performance. In a sensitivity analysis that included halving and doubling the estimate of stage I duration, mortality reduction for annual screening using CA125 and TVS varied by only 2%. Recently, Brown and colleagues estimated that high grade serous cancers spend on average 4 years as *in situ*, stage I and II cancers based on an analysis of pathologic findings in risk-reducing salpingo-oophorectomy specimens in patients with BRCA1 mutations (22). Substituting these more empirically driven estimates of disease duration and limiting analysis to serous EOC did not significantly impact our predictions of screening performance (see Supplementary Table S6).

Our results show that the effectiveness of a multimodal EOC screening strategy is largely dependent on the performance of the first-line test. Substituting the better performing HM for CA125 increased mortality reduction by 12% (from 13% to 25%). Mortality reduction increased to 28% in a sensitivity analysis where the duration of stage I disease was doubled. The impact of improvements in imaging test performance was not as dramatic. In screening strategies where CA125 is used as the first-line test mortality reduction increases only 2% to 3% when TVS is replaced by better performing HI and in a sensitivity analysis where TVS is assumed to have near-perfect sensitivity. When combined with HM as the first-line screen, HI increased mortality reduction by 5% compared with imaging using TVS (26% vs. 21%). We did not evaluate screening strategies that use imaging as a first-line screen because of the reported lower sensitivity of TVS compared with CA125 in early results from both the UKCTOCS and PLCO trials and because HI as a first-line test would likely be prohibitively expensive when used annually. A screening strategy that employs a biomarker as the second-line screen is potentially appealing due to increased sensitivity (23) and low cost. However, we did not model such a strategy out of concern that physicians would be reluctant to proceed to surgery without a confirmatory imaging test.

It is unclear how best to identify biomarkers that perform twice as well as CA125. Despite tremendous effort, no newly discovered EOC biomarkers perform better than CA125 in the critical preclinical phase of the disease (8, 24). Circulating autoantibodies are appealing because of the inherent tendency of the immune system to amplify in response to minute amounts of antigen. Markers that are EOC-specific and not present in unaffected women are also appealing as they lack background signal. This class of markers includes gene fusions and/or translocations that may be discovered using newer deep-sequencing technologies. A candidate gene fusion expressed in roughly 15% of the lethal serous EOC subtype has been recently identified (25). In theory, an imaging test with appropriate performance characteristics could be used as a first-line test. However, cost is a critical issue as the test would be applied to everyone. Our analysis of the impact of cost of the first-line test on the cost-effectiveness of screening suggests that screening using HI as a first-line test at a cost of $750 per test might be feasible if significant mortality reduction could be achieved with screening once every few years. Brown and colleagues estimate that with a screening interval of 24 months it will be necessary to reliably detect tumors no larger than a few millimeters in size to identify 50% of high-grade serous EOC in stage II or earlier. This level of resolution is potentially within reach of newer imaging strategies.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: C.W. Drescher, J.D. Thorpe, M. McIntosh, N. Urban
Development of methodology: C.W. Drescher, J.D. Thorpe, M. McIntosh, S.S. Gambhir, N. Urban
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): N. Urban
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): C.W. Drescher, S. Hawley, J.D. Thorpe, M. McIntosh, S.S. Gambhir
Writing, review, and/or revision of the manuscript: C.W. Drescher, S. Hawley, J.D. Thorpe, S. Particke, M. McIntosh, S.S. Gambhir, N. Urban
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): C.W. Drescher, J.D. Thorpe, S. Particke
Study supervision: C.W. Drescher, S. Particke, N. Urban

Grant Support

This work was supported by the Pacific EOC Research Consortium, Award Number P50 CA083636 from the National Cancer Institute. Support from the NCI U01 CA152637 (to NU and CWD, Chris Li as primary investigator) and the Canary Foundation (to S. Hawley) is also gratefully acknowledged.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 14, 2011; revised May 18, 2012; accepted June 8, 2012; published OnlineFirst July 2, 2012.

References

1. Buys SS, Partridge E, Black A, Johnson CC, Lamerato L, Isaacs C, et al. Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening randomized controlled trial. JAMA 2011;305:2295–303.

2. Menon U, Gentry-Maharaj A, Hallett R, Ryan A, Burnell M, Sharma A, et al. Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative

Trial of Ovarian Cancer Screening (UKCTOCS). Lancet Oncol 2009;10:327–40.
3. Skates S, Troiano R, Knapp RC. Longitudinal CA125 detection of sporadic papillary serous carcinoma of the peritoneum. Int J Gynecol Cancer 2003;13:693–6.
4. Lu KH, Skates S, Bevers TB, Newland W, Moore RG, Leeds L, et al. A prospective U.S. ovarian cancer screening study using the risk of ovarian cancer algorithm (ROCA). 2010 ASCO Annual Meeting J Clin Oncol 2010 (suppl; abstr 5003).
5. Havrilesky LJ, Whitehead CM, Rubatt JM, Cheek RL, Groelke J, He Q, et al. Evaluation of biomarker panels for early stage ovarian cancer detection and monitoring for disease recurrence. Gynecol Oncol 2008;110:374–82.
6. Palmer C, Duan X, Hawley S, Scholler N, Thorpe JD, Sahota RA, et al. Systematic evaluation of candidate blood markers for detecting ovarian cancer. PLoS One 2008;3:e2633.
7. 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;102: 26–38.
8. Cramer DW, Bast RC Jr, Berg CD, Diamandis EP, Godwin AK, Hartge P, et al. Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. Cancer Prev Res 2011;4:365–74.
9. Lutz AM, Willmann JK, Drescher CW, Ray P, Cochran FV, Urban N, et al. Early diagnosis of ovarian carcinoma: is a solution in sight? Radiology 2011;259:329–45.
10. Willmann JK, Kimura RH, Deshpande N, Lutz AM, Cochran JR, Gambhir SS. Targeted contrast-enhanced ultrasound imaging of tumor angiogenesis with contrast microbubbles conjugated to integrin-binding knottin peptides. J Nucl Med 2010;51: 433–40.
11. Willmann JK, Cheng Z, Davis C, Lutz AM, Schipper ML, Nielsen CH, et al. Targeted microbubbles for imaging tumor angiogenesis: assessment of whole-body biodistribution with dynamic micro-PET in mice. Radiology 2008;249:212–9.
12. Havrilesky LJ, Sanders GD, Kulasingam S, Chino JP, Berchuck A, Marks JR, et al. Development of an ovarian cancer screening decision model that incorporates disease heterogeneity: implications for potential mortality reduction. Cancer 2011;117: 545–53.
13. Urban N, McIntosh M, Clarke L, Jacobs I, Karlan B, Anderson G, et al. Socio-economics of ovarian cancer screening. Ovarian Cancer 6. 2nd ed. New York: Oxford University Press; 2002. p. 199–208.
14. Urban N, Drescher C, Etzioni R, Colby C. Use of a stochastic simulation model to identify an efficient protocol for ovarian cancer screening. Control Clin Trials 1997;18:251–70.
15. SEER Cancer Statistics Review. 1973–1991. Bethesda, MD: National Cancer Institute. NIH Publication No. 94-2789; 1994.
16. U.S. Department of Health and Human Services. PHS Centers for Disease Control. Vital statistics of the U.S. 1988. Vol. II. Mortality, Part A, Table 14. Sect. 1, General Mortality, 7; 1991.
17. 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;1:3–16.
18. McIntosh MW, Urban N, Karlan B. Generating longitudinal screening algorithms using novel biomarkers for disease. Cancer Epidemiol Biomarkers Prev 2002;11:159–66.
19. Surveillance, Epidemiology, and End Results (SEER) Program SEER*Stat Database: Incidence - SEER 9 Regs Limited-Use, Nov 2009 Sub (1973–2007) - Linked To County Attributes - Total U.S., 1969–2007 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2010, based on the November 2009 submission. Available from: http://seer.cancer.gov/. Cited: Feb 2012.
20. Appendix: 2003 Statistics on All-listed Procedures in U.S. Hospitals. Available from: http://stats.bls.gov/cpi/home.htm#data. Cited: Feb 2012.
21. Yabroff KR, Lamont EB, Mariotto A, Warren JL, Topor M, Meekins A, et al. Cost of care for elderly cancer patients in the United States. J Natl Cancer Inst 2008;100:630–41.
22. Brown PO, Palmer C. The preclinical natural history of serous ovarian cancer: defining the target for early detection. PLoS Med 2009;6: e1000114.
23. Urban N, Thorpe JD, Bergan LA, Forrest RM, Kampani AV, Scholler N, et al. Potential role of HE4 in multimodal screening for epithelial ovarian cancer. J Natl Cancer Inst 2011;103:1630–4.
24. Zhu CS, Pinsky PF, Cramer DW, Ransohoff DF, Hartge P, Pfeiffer RM, et al. A framework for evaluating biomarkers for early detection: validation of biomarker panels for ovarian cancer. Cancer Prev Res 2011;4:375–83.
25. Salzman J, Marinelli RJ, Wang PL, Green AE, Nielsen JS, Nelson BH, et al. ESRRA-C11orf20 is a recurrent gene fusion in serous ovarian carcinoma. PLoS Biol 2011;9:e1001156.
26. 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;113:775–82.