

A two-step Markov processes approach for parameterization of cancer state-transition models for low- and middle- income countries

Chaitra Gopalappa¹, PhD, Jiachen Guo¹, MS, Prashant Meckoni¹, Buyannemekh Munkhbat¹, Carel Pretorius², PhD, Jeremy Lauer³, PhD, André Ilbawi³, MD, Melanie Bertram³, PhD

¹University of Massachusetts Amherst, Massachusetts, USA

²Avenir Health, Connecticut, USA

³World Health Organization, Geneva, Switzerland

Running head: Parameterization of cancer state transitions for LMICs

Word Count: Abstract- 263; Text- 3963

Corresponding author: Chaitra Gopalappa, chaitrag@umass.edu

Funding: *Financial support for this study was provided by a grant from the World Health Organization.*

The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. The following authors are employed by the sponsor: Jeremy Lauer, André Ilbawi, and Melanie Bertram.

Work conducted at: University of Massachusetts Amherst, Massachusetts, USA; Avenir Health, Connecticut, USA; and World Health Organization, Geneva, Switzerland.

Conference presentation: Work presented at the INFORMS Annual Meeting, November 2014, San Francisco, CA; INFORMS Annual Meeting, November 2017, Nashville, TN;

Abstract

Implementation of organized cancer screening and prevention programmes in high-income countries (HICs) has considerably decreased cancer related incidence and mortality. In low- and middle-income countries (LMICs), screening and early diagnosis programmes are generally unavailable, and a majority of cancers are diagnosed in late stages when survival is very low. Analyzing cost-effectiveness of alternative cancer control programmes and estimating resource needs will help prioritize interventions in LMICs. However, mathematical models of natural cancer onset and progression, for conducting the economic analyses, are predominantly based on populations in HICs because the longitudinal data on screening and diagnoses required for parameterization are unavailable in LMICs. Models currently used for LMICs mostly concentrate on directly calculating the shift in distribution of cancer diagnosis as an evaluation measure of screening.

We present a mathematical methodology for parameterization of natural cancer onset and progression, specifically for LMICs that do not have longitudinal data. This full onset and progression model can help conduct comprehensive analyses of cancer control programmes including cancer screening by considering both the positive impact of screening as well as any adverse consequences such as over-diagnosis and false positive results. The methodology has been applied to breast, cervical, and colorectal cancers for two regions, under the World Health Organization categorization, Eastern Sub-Saharan Africa (AFRE) and Southeast Asia (SEARB). The cancer models have been incorporated into the Spectrum software and interfaced with country-specific demographics data through the demographic projections (DemProj) module and costing

data through the OneHealth tool. These software are open-access and can be used by stakeholders to analyze screening strategies specific to their country of interest.

Introduction

Although overall cancer incidence rates are about twice as high in high-income countries (HIC) as compared to low- and middle-income countries (LMIC), mortality rates are similar between these two country groups.(1-3) These disparities in cancer outcomes are attributed to differences in early detection of cancer through screening and early diagnosis as well as the availability of treatment. For example, in LMIC, most breast cancer cases (e.g. about 70-90% in the African region) get diagnosed in late stages that have low survival rates; this compares to approximately 40% in the United States. (1,4,5) It is thus anticipated that a significant percentage of cancer deaths can be avoided by implementing existing intervention options, such as through organized screening programmes for select cancers including breast cancer and colorectal cancer, improving early diagnosis of cancer, or through vaccinations to prevent cancers caused by infection such as cervical cancer. However, health planners in LMIC are challenged by insufficient economic analyses required for engaging in evidence-based decision-making in prioritization of cancer programmes. Guidelines from HIC are not generalizable given existing resource availability in LMIC.

Conducting economic analyses, i.e., comparing costs and impact of alternative intervention options that considers region-specific resource availabilities and disease burden, helps identify an intervention strategy that is most suitable for the population under analyses. Alternative intervention options that health planners encounter could include (i) different types of screening modalities such as colonoscopy versus faecal immunochemical test for colorectal cancer; (ii) different screening intervals such as mammography every 1 year versus every 2 years to test for breast cancer; or (iii)

combinations of interventions such as frequency of PapSmear to screen for cervical cancer in addition to HPV (human papillomavirus) vaccination to prevent cervical cancer.

Mathematical models of cancer onset and progression, i.e., modeling of state-transitions from healthy to disease-onset, progression through pre-clinical disease, and transition to clinical disease upon diagnoses, are instrumental to conducting such economic analyses. However, there are very few models available that are specific to LMIC, (6) most are based on populations in HIC. (7-13) This is because parameterizing the state-transitions between stages prior to clinical diagnosis is challenging. These parameters cannot be collected through surveillance as they cannot be directly measured (as they occur prior to diagnosis) and thus are mathematically estimated. For high-income countries they have been mostly estimated using longitudinal data from cancer registries, including trends in screening, diagnoses of cancer, and treatment.(7-9,14) In some cases, outcomes of population-based screening studies conducted on small populations have been used to estimate the unknown parameters.(15,16) The pre- and post- screening data provide references for the estimation process. The models used in these estimations are usually Markov-based methodologies that model transitioning of states from healthy, to preclinical disease, to clinical diagnosis, which are governed by rates or probabilities of transitions. They predominantly adopt a technique of systematically varying the unknown pre-diagnosis transition parameters to identify the combination of values that best replicate post-diagnoses population measures, such as incidence, i.e., cancer diagnoses per person-year, and cancer mortality rates, i.e., deaths from cancer per person-year.

Data that are usually available for most LMIC are only cross-sectional data of the outcome measures of incidence and cancer mortality rates.(1,2) There are usually no data on how people are diagnosed, which could vary according to multiple parameters including disease symptoms and access to health care. If we were to apply the technique of systematically varying the unknown parameters, multiple combinations of the unknown parameters could generate similar outcomes as there is no constraint on how people get diagnosed.

In this paper, we present a methodology for parameterizing state-transitions considering data available for LMIC. This method leverages prior work in this area. Specifically, we utilize estimates from HIC populations regarding the progression in preclinical stages of cancer, i.e., after onset of disease and prior to diagnoses, assuming that it remains the same for all populations. We only estimate parameters that likely vary by country, specifically, rates of disease onset per person-year in age group, and diagnostic rates of cancer per person-year by cancer stage. We define *disease onset rates* as the rates of transitioning from healthy to the first (preclinical) stage of disease, which vary across countries based on the population risk of developing disease. We define *diagnostic rates* as the rates of transitioning from pre-clinical to clinical disease state through diagnosis, which may vary across countries based on development of symptoms and access to care. We also estimate the corresponding cross-sectional distribution of the population by disease stage to distribute the base-year population by age and stage in the simulation modeling for the economic analyses.

The methodology was applied to parameterize state-transitions for 2 WHO world regions for 3 types of cancers, breast cancer, cervical cancer, and colorectal cancer.

The parameters were implemented in a simulation model in a user-friendly software package for dissemination to stakeholders for conducting economic analyses of intervention strategies. (17) We present the methodology in the main paper with breast cancer as reference. We present the specific components of the 3 types of cancers and the numerical results in the cancer-specific Appendices.

Methods

The onset of disease and progression of breast cancer is presented in Figure 1. The parameters to estimate are the rates of disease onset θ_a and diagnostic rates $d_{i,a}$, in age a and cancer stage i . Most parameter estimation methods in the literature are based on a Markov processes approach, a stochastic process defined by a *state space* (all possible states the system can occupy) and *transition probabilities or rates* (that transition the system from one *state* to another state over time t). The transitions are governed by the following principles: the state the system transitions to at a specific time-step t is conditional only on the system state at step $t - 1$ and not on any state before that; and, from any state, the probabilities of transition to all other states sum to 1 (i.e., the state space consists of all possible states the system can transition to from any state and the states are mutually exclusive). This framework is convenient to model disease onset, progression, and aging. Age-based health and cancer stages can be represented as states of the Markov process and transition probabilities or rates can be used to transition between these states, e.g., disease onset will be equivalent to transitioning from healthy to disease, and disease progression and regression will be equivalent to transitioning between disease stages. Using the assumption of exponential distribution for transition parameters will maintain the Markov properties.

The commonly used parameter estimation methodology in the literature is to systematically vary the values of the unknown transition parameters to identify ones that provide a good fit to known parameters such as incidence of cancer (i.e., number of yearly transitions to clinical states in Figure 1). However, this method is only feasible when longitudinal data on diagnoses are available such as through small screening studies or cancer registries, as discussed earlier. The majority of LMIC do not have these data. Without this data, multiple combinations of rates of disease onset and diagnosis can generate similar outcomes. Therefore, we propose a two-step methodology that will involve two different Markov processes, i.e., two different state space representations and transition parameters, in each step.

Model 1: We generate a Markov process $Y = \{Y_t; t \geq 0, \Omega, \mathbb{P}, \boldsymbol{\pi}\}$, with a collapsed state space $\Omega = \{[H_a], [U_a], [D_a]\}$ (as in Figure 2), i.e., for each age-group a , $a = \{1, \dots, m\}$, $m = \text{maximum age}$, it collapses all pre-clinical stages into one state U_a and all clinical stages into one state D_a ; H_a represents healthy state; $[.]$ denotes age-vector of states; \mathbb{P} is a matrix of transition probabilities between states for the underlying Markov chain; and $\boldsymbol{\pi}$ is the steady-state distribution vector of the Markov chain. We use this to derive an analytical method to solve for the rates of disease onset θ_a .

Model 2: We generate a Markov process $X = \{X_t; t \geq 0, Z, Q\}$; with the fully extended state space (note the addition of subscript i) $Z = \{[H_a], [U_{i,a}], [D_{i,a}]\}$ representing vectors of states of healthy ($[H_a]$), pre-clinical disease ($[U_{i,a}]$), and clinical disease ($[D_{i,a}]$) at age a and stage i (note: we use a general notation for staging, $i = \{0,1,2,3,4\}$ but actual staging and number of stages could differ by cancer type based on use of

\mathbb{P} =	·											·	
	·											·	
	U ₁					1 - P _{U₁D₁} - P _{U₁M₁}			P _{U₁D₁}			P _{U₁M₁}	
	U ₂						1 - P _{U₂D₂} - P _{U₂M₂}			P _{U₂D₂}			P _{U₂M₂}
	U ₃ *							...			P _{U₃D₃}		P _{U₃M₃}
	·												·
	·												·
	·												·
	D ₁									1 - P _{D₁M₁}			P _{D₁M₁}
	D ₂										1 - P _{D₂M₂}		P _{D₂M₂}
	D ₃ *											...	P _{D₃M₃}
	·											...	
	·											·	
M	1												

Note: empty cells have zero values; '...' refers to additional states corresponding to older age groups that are not shown; *Rows do not add to 1 as not all states are shown as indicated by dotted horizontal and vertical lines

Using the standard definition of risk to rate conversion that assumes that the underlying distributions governing transition probabilities are exponential, the rate of disease onset in age group a can be written as $\theta_a = -\ln(1 - P_{H_a U_a})$, where $P_{H_a U_a}$ is the risk or probability of developing the disease in age a , i.e., an element of \mathbb{P} representing the probability of transitioning from H_a to U_a . Based on the above structure of the Markov process we derived an analytical expression for estimation of $P_{H_a U_a}$ as

$$P_{H_a U_a} = \frac{I_{D_a} c_a - \sum_{k=0}^{a-1} \left(\pi_{H_k} P_{H_k U_k} [\sum_i s_i (1 - e^{-(a-k)\lambda_i}) - \sum_i s_i (1 - e^{-(a-1-k)\lambda_i})] (\prod_{j=k:a+1} e^{-\mu_j}) \right)}{A_a [\sum_i s_i (1 - e^{-\lambda_i})] (e^{-\mu_a}) - I_{D_a} c_a}$$

and developed an iterative process for estimation of $P_{H_a U_a}$ starting with the lowest age.

We present in the Technical Appendix the lemmas and proofs leading to the above

proposition and the iterative process for estimating of $P_{H_a U_a}$ and eventually θ_a , and only outline the final steps of the algorithm in Table 2.

Model 2 ($X = \{X_t; t \geq 0, Z, Q\}$): Estimation of diagnostic rates

For the Markov process X with state space $Z = \{[H_a], [U_{i,a}], [D_{i,a}], M\}$, at age a and stage i ,

$$\rho Q = 0;$$

$$\sum_{k \in S} \rho_k = 1$$

where, ρ is the steady-state distribution vector with element ρ_k for $k \in Z$; and Q is the generator matrix shown below.

	H_a	$U_{0,a}$	$U_{1,a}$	$U_{2,a}$	$U_{3,a}$	$D_{0,a}$	$D_{1,a}$	$D_{2,a}$	$D_{3,a}$	M
H_a	$-\theta_a - \mu_a$	θ_a								μ_a
$U_{0,a}$		$-p_{0,a} - d_{0,a} - \mu_a$	$p_{0,a}$			$d_{0,a}$				μ_a
$U_{1,a}$			$-p_{1,a} - d_{1,a} - \mu_a$	$p_{1,a}$			$d_{1,a}$			μ_a
$U_{2,a}$				$-p_{2,a} - d_{2,a} - \mu_a$	$p_{2,a}$			$d_{2,a}$		μ_a
$U_{3,a}$					$-d_{3,a} - \bar{\mu}_{3,a}$				$d_{3,a}$	$\bar{\mu}_{3,a}$
$D_{0,a}$						$-\bar{\mu}_{0,a}$				$\bar{\mu}_{0,a}$
$D_{1,a}$							$-\bar{\mu}_{1,a}$			$\bar{\mu}_{1,a}$
$D_{2,a}$								$-\bar{\mu}_{2,a}$		$\bar{\mu}_{2,a}$
$D_{3,a}$									$-\bar{\mu}_{3,a}$	$\bar{\mu}_{3,a}$
M	1									-1

*all empty cells =0

The Q matrix consists of transition rates (in per person-year) between stages, where,

$d_{i,a}$ are the diagnostic rates to be estimated for age a , and stage $i = \{0,1,2,3 \dots\}$,

(we use SEER staging for breast cancer and colorectal cancer and TNM for cervical cancer, but for generalization use the notation $i = \{0,1,2,3, \dots\}$)

θ_a are the disease onset rates estimated in Model 1,

$p_{i,a}$ are the progression rates from stage i to $i + 1$; we assume they do not vary by country and use estimates presented in the published literature (see cancer specific Appendix),

μ_a are the disease-free mortalities, and

$\bar{\mu}_{i,a}$ and $\bar{\bar{\mu}}_{i,a}$ are the disease mortalities without treatment and with treatment, respectively, in age a and stage i ; we discuss the estimation in the Technical Appendix, briefly, we estimate the relative increases in these mortalities compared to country-specific cancer-free mortalities (referred to as relative risk) and assume that the relative risks do not vary by populations but the cancer-free mortalities do.

Assumption 1: We set the transition rate from M to $H_1=1$, i.e., we assume births=deaths to convert the model to a *regular* Markov process (if a Markov process is *regular* then the system can transition from any state to any other state in finite number of steps, here replacement of deaths with births makes this possible). The interesting property of *regular* Markov chains include asymptotic convergence to a *unique* steady-state distribution ρ_k , i.e., if one were to assume some arbitrary values for the state-distribution and simulate the process over an extended period of time, then, irrespective of the starting values, the distribution will converge to the same value, approximately; See Remark 1 below.

In Figure 1, s_a , e , and γ represent the rate of cancer screening (defined as inverse of screening interval x participation rate of population), sensitivity of the screening test, and specificity of screening test, respectively. We assume $s_a = 0$, as currently, in the countries of our interest, there is very little screening activity, if any.

Simulation-based optimization for estimation of $d_{i,a}$:

We setup the Markov process as a set of first-order differential equations

$$\rho_{t+1} = \rho_t + \rho_t Q \Delta t \quad (1)$$

that can be simulated over time $t = 1, 2, \dots, T$, where, Δt is a small time-step and T is a large value such that starting with some arbitrary value for ρ_0 , ρ_t asymptotically converges to 0 and $\rho_T \approx \rho_{T-1}$, i.e., the process is at steady state. To estimate the values of $d_{i,a}$ we combine the simulation with an optimization model, with the objective function to minimize the sum of square errors between simulated and actual cancer incidence, i.e., the optimization model is given by,

$$\text{Minimize}_{d_{i,a}} \|\bar{\mathbf{I}} - \mathbf{I}\|_2, \quad d_{i,a} \geq 0; a = 1, \dots, m; i \in \{0, 1, 2, 3, 4\} \quad (2)$$

$\mathbf{I} = [I_1, I_2, \dots, I_m]$ are the actual cancer incidence per person-year in age-group $a = 1, 2, \dots, m$, which we assume are known (see Table 1), and

$\bar{\mathbf{I}} = [\bar{I}_1, \bar{I}_2, \dots, \bar{I}_m]$ are the simulated cancer incidence at steady-state (i.e., at $T=500$ in (1)), and is a function of the decision variable $d_{i,a}$, i.e., $\bar{\mathbf{I}} = f(d_{i,a})$, and

$\|\cdot\|_2$ is the Euclidean norm.

We used a simulation-based optimization methodology to solve this problem. In regular optimization models, the analytical form of the objective function, i.e., a regression equation with the decision variables as the independent parameters are known. In the above problem, the objective function is $\|\bar{\mathbf{I}} - \mathbf{I}\|_2 = f(d_{i,a})$, where $f(\cdot)$ is some function on $d_{i,a}$, but the analytical form of $f(d_{i,a})$ is not known. However, for any specific value of $d_{i,a}$ the value of $\|\bar{\mathbf{I}} - \mathbf{I}\|_2$ can be estimated using the simulation model discussed

above. The problem was coded in MATLAB, version R2014a. We combined the simulation algorithm with an in-built optimization solution algorithm (lsqcurvefit function in the MATLAB optimization toolbox), that searches through the decision space to find $d_{i,a}$ that minimizes the error, i.e., $\|\bar{I} - I\|_2 \approx 0$. The simulation-based optimization algorithm is summarized in Table 3. We present in the Technical Appendix the conditions for *global* optimality, i.e., the conditions that guarantee a *unique* optimal solution to the problem.

Remark 1: We normalize the steady-state distributions by age, i.e., we estimate $\bar{\rho}_{k_a} = \frac{\rho_{k_a}}{M_a}$; $M_a = \rho_{H_a} + \sum_i \rho_{U_{(i,a)}} + \sum_i \rho_{D_{(i,a)}}$, $k_a \in \{H_a, U_{i,a}, D_{i,a}, M\}$; such that $\bar{\rho}_{H_a} + \sum_i \bar{\rho}_{U_{(i,a)}} + \sum_i \bar{\rho}_{D_{(i,a)}} = 1$, for every age-group a . In application to a country, we use the actual age distribution of the population from the demographics data and use $\bar{\rho}_{k_a}$ to distribute each age-group of base-year population into disease stages. We then use actual demographics projections to simulate future years, and thus the state distributions dynamically change over time. This method will minimize estimation errors that might result from Assumption 1 (see Technical Appendix for sensitivity analyses on Assumption 1).

Application and software for model dissemination:

We applied the methodology to parameterize the state-transitions for 3 types of cancers, breast, cervical, and colorectal cancers, for 2 world regions, Eastern Sub-Saharan Africa (AFRE) and Southeast Asia (SEARB). We discuss, in the cancer-specific Appendix, further details of the calibration specific to each cancer. Using these state-transitions, we implemented a cancer simulation of the form in Model 2 in the non-

communicable diseases (NCD) module of the Spectrum software package.(18) The cancer module is interfaced with the demographic projections module (DemProj) in Spectrum, which predicts future demographic projections for individual countries, and the One-Health tool (19) that does ingredients-based costing of interventions. These models were used for economic evaluations of interventions under the WHO-CHOICE (CHOosing Interventions that are Cost-Effective)(22,23) analyses, for updating the Appendix 3 of the Global Action Plan for Non-Communicable Diseases (24), which provides United Nations Member States a list of generalized cost-effective policy options for achieving the health goals of the Global Action Plan. Details of the analyses and results are presented in WHO Technical Briefings and WHO Discussion Papers, (25,26) accompanying documents to the Appendix 3 of the Global Action Plan. Spectrum and One-Health are publicly available software packages that can be downloaded from Avenir Health website. (17) This user-friendly software package can be used by health planners to further conduct country-specific analyses for identifying a strategy most suitable for the population or estimating resources needed for a suitable intervention strategy. Users also have an option of modifying parameter inputs, including transition rates and screening and treatment options and their sensitivity and specificity values, to more closely represent the population and resource availabilities and for conducting sensitivity analyses.

Results

To verify the parameterized results for the natural history model, we compared the simulated cancer incidence with rates from GLOBOCAN, which were originally used in the parameterization process (see cancer Appendices). To validate the model, we

compared results with independent data not used in model parameterization, e.g., total mortality rates among persons with breast cancer from GLOBOCAN, or HPV prevalence from clinical studies for cervical cancer model. The results were generally within acceptable values, the details are presented in the cancer Appendices.

Discussion

We present a two-step methodology for parameterizing a Markov processes model for cancer onset and progression. The methodology is general and intended for estimating national averages in countries where longitudinal cancer registry data are not available, which impedes construction of a more detailed model. Health planners or other stakeholders can utilize the models in the Spectrum software to analyze impact and costs of alternative intervention scenarios specific to the region of their interest, selecting between cancer control interventions for prevention or screening for early diagnosis. Alternative intervention options that health planners encounter could include (i) different types of screening modalities such as colonoscopy versus fecal immunochemical test for colorectal cancer; (ii) different screening intervals such as mammography every 1 year versus every 2 years to test for breast cancer; or (iii) combinations of interventions such as frequency of PapSmear to screen for cervical cancer in addition to HPV (human papillomavirus) vaccination to prevent cervical cancer.

The methodology is subject to certain limitations. To meet the mathematical properties required for estimation of parameters, i.e., steady state, we made certain assumptions such as births are equal to deaths to make the Markov chain *regular*. We also assumed that the Markov processes are *stationary*, i.e., the transition parameters

do not change over time, including cross-sectional measures of incidence and cancer stage distribution at diagnosis. Use of cross-sectional data on incidence and assumptions of its stationarity over time ignores effects of temporal changes in competing risks of mortality as well as the effects of temporal changes in cancer-specific parameters that affect incidence, progression, access to care and other relevant variables. Problems of cohort-specific effects are not unique to cancers, or to LMICs, but they remain an unquantifiable source of uncertainty for models of this kind. More specifically, use of cross-sectional data generates 2 types of limitations. First, it does not capture evolutionary changes in cancer, i.e., changes in risk of disease (onset rates) and disease severity (progression rates) that are anticipated to occur due to environmental or behavioral changes, e.g., changes in diet or delays in childbearing for breast cancer.(20,21) Second, it also ignores the effects of temporal changes in population demographics including changes in competing risks of mortality from other diseases, such as changes in maternal and infant mortality, or mortalities from infectious diseases. Therefore, as data or more reliable predictions on population changes become available, analyses in this study should be redone. There are no reliable predictions on evolutionary changes in cancer risk. To make best use of available projections on population size estimates, the cancer models in Spectrum are interfaced with the Demographic Projections (DemProj) model, (17) a model that estimates future population sizes of a country or region by age and sex and is updated frequently with the latest estimates for future trends in fertility rates and overall mortality rates supplied by the Population Division of the United Nations. To effectively connect our static results with these dynamic population projections the results from this study

are normalized by age. During simulations of cancer incidence over time, the normalized outputs interface with the future projections for demographic distributions of a country in the DemProj model (see Technical Appendix). However, this does not consider the changes in trends in disease parameters and thus does not replace the need for re-conducting this analyses over time as more data become available. Any cost-effectiveness analyses conducted using results from this study will also be subject to these limitations and should be used with caution.

In application of the methodology to parameterize 2 regions, AFRE and SEARB, we only considered heterogeneity by age and did not consider any other population characteristics and differences across countries. Causal factors for differences across countries in the risk of disease could be multiple, including diet, alcohol and tobacco consumption, competing diseases, or genetic. We did not explicitly consider these factors, but the estimated disease onset parameter might include these factors collectively. For persons who were diagnosed, we did not explicitly model recurrence of disease, we only applied an average stage-specific rate of survival. We assumed that for persons with the disease, progression rates in preclinical stages do not vary by populations. Diagnostic rates could be a function of development of symptoms, delays in correct diagnosis, and/or access to care. We do not explicitly consider these factors in the estimation process. The transition parameters were grouped according to general cancer types. Different cancer subtypes, such as hormone receptor positive breast cancers, were not considered in this study. This model thus assumes that there is not significant heterogeneity in the cancer subtypes between different populations.

The data used were gross average estimates, as the scope of our work was to generate aggregated estimates for the WHO-CHOICE and WHO Global Action Plan.(24) Nonetheless, we believe these applications, to a certain extent, helped in the validation of the new methodology presented here. Application of the model to individual countries should consider more country-specific data inputs as available, and conduct sensitivity analyses around the input parameters for evaluating the impact of parameter variabilities on program decisions. Validation of the model were restricted to comparison of a few parameters, not used in estimation, but aggregated at regional-levels. Future work could include more extensive validation by applying the methodology to populations with known parameter estimates.

The model presented here complements the current models in the GLOBOCAN,(1) WHO-CHOICE (CHOosing Interventions that are Cost-Effective),(22,23) and OneHealth(19) projects. It uses incidence estimates of diagnosed cases of cancer from GLOBOCAN to construct the disease onset and progression model. It serves as an 'impact module' for analyzing the impact of alternative interventions under the WHO-CHOICE work. The Spectrum package also contains models for other non-communicable diseases, which enables a health systems approach to analyses of multiple diseases. These together will facilitate health planners to evaluate cancer-intervention decisions in a context specific to a country, i.e., in addition to impacts and costs of interventions for cancer, consideration of competing health priorities, availability of specific treatment options or other intervention techniques, and infrastructure availabilities. The methodology developed for model parameterization can be extended to other types of cancers.

Acknowledgements: We acknowledge the following persons for their work in developing this manuscript. Michael McGrath, UMass Amherst, provided technical help in MATLAB, in automating the mathematical model for application to multiple countries. Elric Werst, Avenir Health, provided software assistance in integrating the model into Spectrum. Vijeta Deshpande helped generate part of the results for the cancer Appendices. Rachel Sanders and Nadia Carvalho, Avenir Health, helped interface the impact module in Spectrum with the OneHealth costing module. We also acknowledge the comments from the reviewers and the journal editor, which helped significantly improve the presentation of the work.

References

- (1) Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: <http://globocan.iarc.fr>, accessed on 2/22/2016.
- (2) Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015 March 01;65(2):87-108.
- (3) Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011 April 01;61(2):69-90.
- (4) Harford JB, Otero IV, Anderson BO, Cazap E, Gradishar WJ, Gralow JR, et al. Problem solving for breast health care delivery in low and middle resource countries (LMCs): consensus statement from the Breast Health Global Initiative. *Breast* 2011 April 01;20 Suppl 2:S20-9.
- (5) Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol* 2006 May 10;24(14):2137-2150.
- (6) Zelle SG, Baltussen RM. Economic analyses of breast cancer control in low- and middle-income countries: a systematic review. *Syst Rev* 2013 April 08;2:20-4053.
- (7) Mandelblatt J, Schechter CB, Lawrence W, Yi B, Cullen J. The SPECTRUM population model of the impact of screening and treatment on U.S. breast cancer trends from 1975 to 2000: principles and practice of the model methods. *J Natl Cancer Inst Monogr* 2006;(36)(36):47-55.
- (8) Tan SY, van Oortmarssen GJ, de Koning HJ, Boer R, Habbema JD. The MISCAN-Fadia continuous tumor growth model for breast cancer. *J Natl Cancer Inst Monogr* 2006;(36)(36):56-65.
- (9) Tan KH, Simonella L, Wee HL, Roellin A, Lim YW, Lim WY, et al. Quantifying the natural history of breast cancer. *Br J Cancer* 2013 October 15;109(8):2035-2043.
- (10) Goldie SJ, Grima D, Kohli M, Wright TC, Weinstein M, Franco E. A comprehensive natural history model of HPV infection and cervical cancer to estimate the clinical impact of a prophylactic HPV-16/18 vaccine. *Int J Cancer* 2003 October 10;106(6):896-904.
- (11) Chen CD, Yen MF, Wang WM, Wong JM, Chen TH. A case-cohort study for the disease natural history of adenoma-carcinoma and de novo carcinoma and surveillance of colon and rectum after polypectomy: implication for efficacy of colonoscopy. *Br J Cancer* 2003 June 16;88(12):1866-1873.
- (12) Leshno M, Halpern Z, Arber N. Cost-effectiveness of colorectal cancer screening in the average risk population. *Health Care Manag Sci* 2003 August 01;6(3):165-174.

- (13) Frazier AL, Colditz GA, Fuchs CS, Kuntz KM. Cost-effectiveness of screening for colorectal cancer in the general population. JAMA 2000 October 18;284(15):1954-1961.
- (14) Fryback DG, Stout NK, Rosenberg MA, Trentham-Dietz A, Kuruchittham V, Remington PL. The Wisconsin Breast Cancer Epidemiology Simulation Model. J Natl Cancer Inst Monogr 2006;(36)(36):37-47.
- (15) Duffy SW, Day NE, Tabar L, Chen HH, Smith TC. Markov models of breast tumor progression: some age-specific results. J Natl Cancer Inst Monogr 1997;(22)(22):93-97.
- (16) Tabar L, Vitak B, Chen HH, Prevost TC, Duffy SW. Update of the Swedish Two-County Trial of breast cancer screening: histologic grade-specific and age-specific results. Swiss Surg 1999;5(5):199-204.
- (17) Spectrum software, Avenir Health, <http://www.avenirhealth.org/software-spectrummodels.php#demproj> last accessed 12-18-2017
- (18) Stover J, Andreev K, Slaymaker E, Gopalappa C, Sabin K, Velasquez C, et al. Updates to the spectrum model to estimate key HIV indicators for adults and children. AIDS 2014 November 01;28 Suppl 4:S427-34.
- (19) World Health Organization, OneHealth Tool: Supporting integrated strategic health planning, costing and health impact analysis, World Health Organization, 2013 http://www.who.int/choice/onehealthtool/OneHealth_Tool_Supporting_integrated_strategic_health_planning.pdf?ua=1 (last accessed January 2016).
- (20) Bray F, Jemal A, Grey N, Ferlay J, Forman D. Global cancer transitions according to the Human Development Index (2008-2030): a population-based study. Lancet Oncol 2012 August 01;13(8):790-801.
- (21) Porter P. "Westernizing" women's risks? Breast cancer in lower-income countries. N Engl J Med 2008 January 17;358(3):213-216.
- (22) List of publications that use World Health Organization- Choosing Interventions that are Cost-effective (WHO-CHOICE) for economic evaluation, WHO, http://www.who.int/choice/documents/economic_evaluation/en/ (last accessed 7-19-2016).
- (23) World Health Organization, WHO-CHOICE Frequently asked questions, World Health Organization, 2014 http://www.who.int/choice/cost-effectiveness/CHOICE_FAQ.pdf?ua=1 (last accessed January 2016).
- (24) World Health Organization, Preparation for the third High-level Meeting of the General Assembly on the Prevention and Control of Non-communicable Diseases, to be held in 2018, December 2016, pg 21-22 http://apps.who.int/gb/ebwha/pdf_files/EB140/B140_27-en.pdf?ua=1 <http://www.who.int/ncds/governance/appendix3-update/en/> <http://www.who.int/ncds/governance/appendix3-update-discussion-paper/en/> [last accessed- 7-11-2017].
- (25) World Health Organization, Cancer Intervention Technical Briefing, <http://www.who.int/ncds/governance/appendix3->

[update/en/ http://www.who.int/ncds/governance/cancers.pdf?ua=1](http://www.who.int/ncds/governance/cancers.pdf?ua=1) [last accessed- 7-11-2017].

(26) World Health Organization, Draft Updated Appendix 3 of the WHO Global NCD Action Plan 2013, pg 25-26. <http://www.who.int/ncds/governance/discussion-paper-updating-appendix3-25july2016-EN.pdf?ua=1> [last accessed- 12-25-2016].