e-Health – For Continuity of Care
C. Lovis et al. (Eds.)

doi:10.3233/978-1-61499-432-9-288

Estimating the Impact of Prevention Action: A simulation Model of Cervical Cancer Progression

Michal ROSEN-ZVI[a1], Lavi SHPIGELMAN[a], Alan KALTON[b], Omer WEISSBROD[a], Saheed AKINDEINDE[b], Soren BENEFELDT[c], Andrew BENTLEY[d], Terry EVERETT[c], Joseph JAJINSKIJI[e], Emmanuel KWEYU[f], Chalapathy NETI[e], Joe SAAB[c], Osamuyimen STEWART[b], Malcolm WARD[c], Guo Tong XIE[g]
[a]*IBM Research – Haifa*
[b]*IBM Research –Africa*
[c]*IBM Services*
[d]*IBM Software Group*
[e]*IBM Research – Watson*
[f]*Stathmore University, Kenya*
[g]*IBM Research - China*

Abstract. Cervical cancer is one of the highest occurring cancers for women in East Africa. Many studies have shown that disease occurrences and particularly the number of deaths due to the disease can be reduced significantly by screening and vaccination. East Africa and Kenya in particular are undergoing change and taking actions to reduce disease levels. However, up until today disease level in the different districts in Kenya is not known nor what be the prevalence of disease when prevention actions take place. In this paper we propose a novel Bayesian model for estimating disease levels based on available partial reports and demographic information. The result is a simulation engine that provides estimations of the impact of various potential prevention actions.

Keywords. Disease prevalence, Bayesian Model, Cervical Caner

Introduction

Cervical cancer remains one of the strongest threats to female reproductive health; it is the second most common cancer worldwide and in the Sub-Saharan African region, and is the leading cause of cancer related death among women in developing countries. The human papillomavirus (HPV) is known to be the causative agent of a vast majority of cervical cancers. Early detection of the virus by screening and removal of the precancerous cells, as well as vaccination against (most of) the HPV strains that cause the cancer have reduced the level of disease in more developed countries. Current screening tests include the Pap test that explores changes in the cells of the cervix, the HPV test that detects the existence of the high risk strains of the HPV virus in the cells, visual inspection with acetic acid (VIA), and visual inspection with Lugol's iodine

[1] Corresponding author.

(VILI). Not only is the cost of VIA and VILI significantly lower than the PAP test and HPV test but these visual inspection methods also allow for treating of any detected abnormal cells during the same visit as the screening. These 'see and treat' tests are a better fit for countries with fewer resources and have become common in Kenya in recent years.

Vaccinations, screening tests and appropriate treatment have become part of national policies and guidelines formulation in many countries. The Kenyan Ministry of Health adopted guidelines developed by the Department of Reproductive Health and partners as a standard approach to be used by all cadres of health care workers in the country; these guidelines address screening, diagnosis and palliative care for cervical cancer patients. However, they have been only partially adopted and with a large variation between different districts in Kenya due to constraint resources. Moreover the information on such actions is not centrally gathered and it is a challenge to estimate the current and expected disease levels. We propose a system that is based on an innovative Bayesian model [1] that utilizes the unique freely available dataset of Kenya called Open Kenya Data [2] and provides district level estimates of disease prevalence.

1. Methods

In this work we leverage existing models for cervical cancer progression that have been used to estimate disease level and cost effectiveness of vaccination and screening, and extend the models to include demographic information that in turn enables more accurate estimates of current disease level. We infer the dependencies and values of the parameters and random variables in the Bayesian model by performing a meta-analysis study. To the best of our knowledge this is the first time a Bayesian model is introduced to tie between rich demographic information and disease status using meta-analysis in general and the first to be used for modeling Cervical Cancer prevalence and progression.

Bayesian models are statistical models represented by acyclic graphs that capture dependencies between random variables. In this paper we used the convention that random variables are represented by circles, arrows stand for dependency and shaded circles stand for random variables for which the value is considered observed. In this paper we propose a model that stands for different characteristics of a woman in Kenya, her particular district, whether she lives in rural or urban area, likelihood of her having cervical cancer and so on. The random variables in this model are those characteristic and take values from a finite set of values (also the age variable is categorized into a few age groups). Figure 1 illustrates the model.

A number of Bayesian models have been introduced to model the progression from HPV infection to a cervical cancer. One of the most commonly used models is the Markov model introduced by Mayers et al[3] that captures the different stages from infectious by the virus where the woman state is 'well' to the presence of a Low-grade Squamous Intraepithelial Lesion ('LSIL') (cervical intraepithelial lesion 1). Such state might progress to High-grade SIL ('HSIL') (cervical intraepithelial lesion 2-3, including carcinoma in situ). The woman might progress to have local than regional than distant cancer. In Figure 1, the insert illustrates the model for such natural progression. The model we propose captures the likelihood for a woman to be screened, and treated given that pre-cancerous cells or cancer is detected.

Figure 1: Dynamic Bayesian model of cervical cancer progression and monitoring

The likelihood for a screening depends on availability and accessibility of the screening as represented by the information whether the woman has access to a healthcare facility nearby. The likelihood for a woman living in a city to have a healthcare facility nearby is larger than those who leave in rural areas. This is one way by which the demographic information plays a role in the estimations of disease prevalence. Similarly, as vaccination typically is offered to girls in school, as according to guidelines the vaccine is most effective while administered before getting engaged in sexual activity, information regarding school attendance can serve as a predictor of whether or not a woman is vaccinated, and thus the likelihood of contracting the disease. This information is also captured in our proposed model, Figure 1.

Dynamic Bayesian models [4] allowing the state space to be represented in factored form, instead of as a single discrete random variable. A given hypothetical state of a woman is characterized by her age group, where she lives (in rural or urban area), whether she is infected by HPV, if she has a cancer, and if yes in what state and so on. The development of the state of the woman over the years is depicted by a dynamic Bayesian model, where in Figure 1 gray arrows depict dependencies between different variables within the same timeframe and green arrows depict the dependency between the states in one year to the following one.

We used this innovative dynamic model by extracting real information from Open Kenya data with respect to the demographic information on women in Kenya. We learned dependencies between the random variables by exploiting standard meta-analysis. The model than can serve to simulate different scenarios such as what happens in terms of disease prevalence five and ten years ahead if a portion of the woman are vaccinated and / or screening policy is applied.

Finally, we used existing references and meta-analysis method [5] for evaluating the Bayesian model results with respect to current disease prevalence and prevention actions. Search of PUBMED with the keyword “cervical cancer AND Kenya” produced 99 related articles on which we applied our inclusion and exclusion strategy. In addition we performed a few discussions with local nonprofit organizations and obtained links to reports of disease status and associated studies that are not covered by PUBMED, The initial strategy was a quick review of the abstracts/reports for a description of the study on desired data of interest. The inclusion or exclusion of published articles where only abstracts were accessible was based solely on whether

data of interest are reported in the abstracts or not. This way, the initial search result list was filtered to 24 articles. A detailed review of these articles (reading the entire paper) revealed that some data of interest are only partially reported.

2. Results

The meta-analysis study yielded a few but significant references such as [6],[7],[8]. These reports where used to confirm the Bayesian model analysis at the country level. The estimates we have performed where per district. There are close to 13 Million women at age 14 and higher out of a total population of around 44 Million.

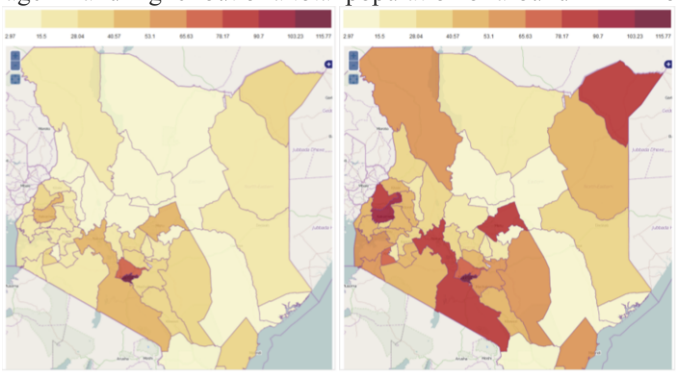

Figure 2: Nunber of women (x1000) aged 35-40 (left) and 9-12 (right) per distrcit

We used the Bayesian model to simulate an estimate of mortality rate for 35-40 year old women in 10 years in three different conditions; status quo, screening and treating as necessary 90% of all women in current year and the same but only for those living near a health facility. Figure 3 shows these scenarios. The redder the color, the higher the number.

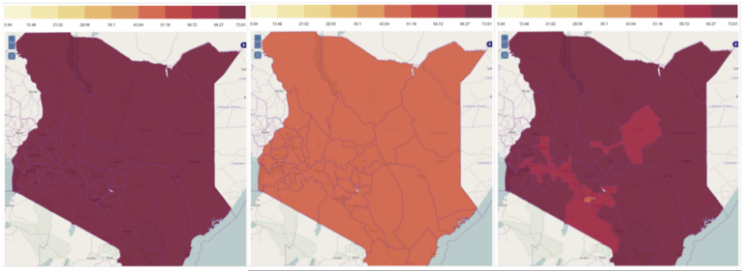

Figure 3 mortality rate per 100K women aged 35-40 in 10 years. Left: status quo. Middle: screen & treat 90% this year. Right: screen & treat 90% who live near health facility

Table 1 compares cost effectiveness of the mentioned scenarios (assuming pricing similar to Goldie et al [9]). While vaccination of all 9-12 yo is relatively expensive and has no immediate effects, it saves the most lives and life-years in the long run.

Table 1: Cost-effectiveness of simulated scenarios

solution	cost ($K)	cost per life-year-saved ($ / yr)	healthcare savings in later 10 years (dollars)	lives saved (10-yr pd) (women)	lives saved (30-yr pd) (women)	life-years saved (30-yr pd) (years)	life-years saved (10-yr pd) (years)	cancer incidence rate in 10 years (women per 100K)	cancer incidence rate in 30 years (women per 100K)	population (1000 women)
9-12 year olds: status quo	0	0	0	0	0	0	0	30.29	80.37	2090.6
9-12 year olds: vaccinate	21845.98	368.54	13786.55	262.77	6526.99	59276.55	624.52	18.57	50.25	2090.6
35-40 year olds: status quo	0	0	0	0	0	0	0	103.32	158.13	1194.74
35-40 year olds: screen and treat all	15522.75	221.15	25130.88	1439.6	4817.01	70190.06	5047.17	78.43	154.38	1194.74
35-40 year olds: screen and treat near HF	2825.78	134.31	17542.26	435.39	1436.44	21039.24	1528.95	95.89	157.06	1194.74

3. Discussion

Our proposed system can serve agencies in making informed decisions with respect to planned programs to launch. At the moment, health records in Kenya are very limited, and yet they indicate a high prevalence of cervical cancer, low level of awareness of the disease and limited access to screening among women in Kenya. A coordinated effort is performed between different agencies to raise the awareness for the disease, increase level of screenings, and apply vaccination programs. The system we developed can easily integrate further information, at the form of evidence in the Bayesian model and to provide more accurate results when more information is collected with respect to disease level. We are working with these forces and aim at supporting further coordinating the effort that hopefully will lead to a significant decrease of the number of women suffering from cervical cancer in Kenya. This system can be easily extended for usage in other disease areas.

References

[1] J Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kauffman , San Francisco (1988)
[2] Kenya open Data, https://opendata.go.ke/ accessed on 14th of August 2013
[3] ER Myers, DC McCrory, K Nanda, L Bastian , and D Matchar (2000). Mathematical Model for the Natural History of Human Papillomavirus Infection and Cervical Carcinogenesis. Am. J. Epidemiol. 151 (12): 1158-1171
[4] KP Murphy, Dynamic Bayesian Networks: Representation, Inference and Learning. Dissertation. Univ. of California (2002)
[5] D F Stroup, J A Berlin, S C Morton, et al. (2000). Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA : the journal of the American Medical Association, 283(15), 2008-2012
[6] J S Smith, S Moses, M G Hudgens, et al. Increased Risk of HIV Acquisition among Kenyan Men with Human Papillomavirus Infection. J Infect Dis 2010; 201(11):1677–1685
[7] "Kenya first country to protect girls against cervical cancer with GAVI support, GAVI Alliance", 14 May 2013, (http://www.gavialliance.org/library/news/press-releases/2013/kenya-first-country-to-protect-girls-against-cervical-cancer-with-gavi-support/#sthash.6XwVgnka.dpuf)
[8] WHO/ICO Information Centre on HPV and Cervical Cancer (HPV Information Centre). Human Papillomavirus and Related Cancers in World. Summary Report 2010.
[9] Goldie, S. J., Kuhn, L., Denny, L., Mmed, O., Pollack, A., & Wright, T. C. Policy Analysis of Cervical Cancer Screening Strategies in Low-Resource Settings, JAMA 2010, *285*(24), 3107–3116.