Cancer Epidemiology 53 (2018) 81–92

Contents lists available at ScienceDirect

Cancer Epidemiology

journal homepage: www.elsevier.com/locate/canep

# Population risk factors for late-stage presentation of cervical cancer in sub-Saharan Africa

Tessa S. Stewart[a,*], Jennifer Moodley[b,c,d], Fiona M. Walter[e]

[a] The University of Cambridge School of Clinical Medicine, Addenbrookes Hospital, Hills Road, Cambridge CB2 0SP, United Kingdom
[b] Women's Health Research Unit, School of Public Health and Family Medicine, University of Cape Town, Anzio Road, Observatory, Cape Town 7925, South Africa
[c] Cancer Research Initiative, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, Cape Town 7925, South Africa
[d] SAMRC Gynaecology Cancer Research Centre, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa
[e] The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, United Kingdom

ARTICLE INFO

*Keywords:*
Uterine cervical neoplasms
Africa south of the Sahara
Incidence
Prevalence
Risk factors
Condoms
Rural population
HIV infections
Socioeconomic factors

ABSTRACT

*Background:* Cervical cancer is the most prevalent malignancy in sub-Saharan Africa (SSA) with many women only seeking professional help when they are experiencing symptoms, implying late-stage malignancy and higher mortality rates. This ecological study assesses population-level exposures of SSA women to the numerous risk factors for HPV infection and cervical cancer, against late-stage presentation of cervical cancer.

*Materials and method:* A literature review revealed the relevant risk factors in SSA. Open-access databases were mined for variables closely representing each risk factor. A proxy for late-stage presentation was used (ratio of incidence-to-mortality, IMR), and gathered from IARC's GLOBOCAN 2012 database. Variables showing significant correlation to the IMR were used in stepwise multiple regression to quantify their effect on the IMR.

*Results:* Countries with high cervical cancer mortality rates relative to their incidence have an IMR nearer one, suggesting a larger proportion of late-stage presentation. Western Africa had the lowest median IMR (1.463), followed by Eastern Africa (IMR = 1.595) and Central Africa (IMR = 1.675), whereas Southern Africa had the highest median IMR (1.761). Variables selected for the final model explain 65.2% of changes seen in the IMR. Significant predictors of IMR were *GDP* (coefficient = $2.189 \times 10^{-6}$, p = 0.064), *HIV infection* ($-1.936 \times 10^{-3}$, p = 0.095), *not using a condom* ($-1.347 \times 10^{-3}$, p = 0.013), *high parity* ($-1.744 \times 10^{-2}$, p = 0.008), and *no formal education* ($-1.311 \times 10^{-3}$, $p < 0.001$).

*Conclusion:* Using an IMR enables identification of factors predicting late-stage cervical cancer in SSA including: *GDP, HIV infection, not using a condom, high parity* and *no formal education.*

## 1. Introduction

Cervical cancer is the most prevalent malignancy among women in the developing world. It is also the most common AIDS-related cancer in women [1,2] and is responsible for the majority of cancer-related deaths within the four regions of sub-Saharan Africa (SSA) (Eastern, Central, Southern and Western) [3]. Of these, Eastern Africa reportedly has the highest incidence of cervical cancer worldwide [4]. These ominous data seem at odds with the fact that cervical cancer is the most preventable form of cancer through vaccination, screening and treatment of pre-cancerous lesions [2,5]. Central to the development of cervical cancer is infection by the sexually transmitted mucosal human papillomavirus (HPV) [6,7]. Worldwide, HPV prevalence in women with normal cervical cytology is approximately 12%, whereas in SSA the prevalence averages at 24%, ranging between 17.4% and 33.6% [8]. To date, the most effective cervical cancer control method has been screening programmes that detect women at increased risk of developing cervical cancer, followed by treatment of the pre-cancerous lesions prior to malignant transformation.

The high burden of cervical cancer in SSA is partly explained because cytology-based screening is prohibitively difficult in those countries lacking adequate health infrastructure, resources and expertise. There are currently very few SSA countries that possess both the capacity and expertise to perform cytology-based screening in a primary care setting nationwide [2,9]. In addition, due to competing demands on the healthcare system from HIV and tuberculosis, health system challenges and lack of human resources, the coverage achieved by cytology screening programmes has varied between less than 10%

*Abbreviations:* SSA, sub-Saharan Africa; IMR, incidence-to-mortality ratio; HPV, human papillomavirus; LMIC, low-and middle-income countries; IARC, International Agency for Research on Cancer; DHS, USAID's Demographic and Health Survey; GDP, gross domestic product

* Corresponding author.
*E-mail addresses:* tss32@cam.ac.uk (T.S. Stewart), Jennifer.Moodley@uct.ac.za (J. Moodley), fmw22@medschl.cam.ac.uk (F.M. Walter).

https://doi.org/10.1016/j.canep.2018.01.014
Received 19 September 2017; Received in revised form 18 January 2018; Accepted 22 January 2018
1877-7821/ 

and 50% [10–14].

Consequently, many women only seek help from healthcare professionals once they are experiencing gynaecological symptoms (unexpected vaginal bleeding, foul-smelling vaginal discharge, haematuria and abdominal pain) [13,15,16]. This symptomatic presentation means the malignancy is already at an advanced stage, resulting in a poor prognosis and increased mortality [17–19]. There are numerous risk factors, both physiological and socio-cultural, that make SSA women particularly vulnerable to developing cervical cancer. Ginsburg et al. [20] demonstrated that factors including where a woman lives, or if she is affected by poverty, largely determine whether she will develop cervical cancer and her chances of survival within low- and middle-income countries (LMICs).

In this ecological study we assessed the relationship between population-level exposure to established risk factors and potential regional differences with late-stage cervical cancer presentation.

## 2. Methods

### 2.1. SSA countries and their cervical cancer prevention programmes

Sub-Saharan Africa are those 51 countries predominantly located south of the Sahara desert, and divided into Eastern, Southern, Central and Western Africa [22]. Owing to a lack of data, five countries were omitted from the analysis: Sao Tome and Principe, Mayotte, Reunion, Seychelles, and Saint Helena. The regional groupings, adult female population size [23], presence and type of cervical cancer prevention programme and female population coverage achieved in each country, can be seen in Table 1.

### 2.2. Computed variables

As the stage of cervical malignancy at presentation is not routinely reported for all SSA countries, using a ratio of the incidence-to-mortality (IMR) [24,25] may provide a means of estimating the level of late-stage presentation for each country. Consequently, a ratio closer to one implies a late stage of presentation, as both the incidence and mortality values are similar i.e. a larger proportion of the population is dying from cervical cancer than is expected relative to the incidence rates. Conversely, countries yielding a higher IMR are more likely to have fewer late-stage cervical cancer presentations. The inverse of this ratio, mortality-to-incidence, has been shown as a reliable proxy for 1-survival and an important measure of completeness of cancer registry data and the quality of cancer care in both a single country [26] and in multiple countries [25]. Cervical cancer data was derived from IARC's GLOBOCAN 2012 database [4] and included age-standardised incidence and mortality rates per 100,000 of the female population. The IMR for each country is shown in Table 1.

### 2.3. Risk factors and data sources

A literature review of risk factors for HPV infection and the development of cervical cancer in SSA was undertaken. Population-level data for each risk factor from each SSA country were collected using open-source databanks. The risk factors and the variables used, where possible, are shown in Table 2. The World Bank Databank [23] provided the demographic, female HIV and smoking prevalence, undernourishment and parity data. UNESCO Institute for Statistics: Data Centre [32] provided education-related data, whilst sexual behaviours, marital status, contraception, STIs and additional education data was obtained from USAID's Demographic and Health Surveys (DHS) Program database [33]. Wherever possible, data was taken from 2012 in accordance with the GLOBOCAN cervical cancer data. In instances where 2012 data was not available, the nearest year in which all countries had available data was selected.

### 2.4. Variable selection

An initial culling of variables occurred after Table 2 was compiled, when risk factors for which there were significant missing data or no attainable variables were excluded. Demographic and Health Survey data were not available for the following regions: Central Africa (Angola, Central African Republic, Chad), Eastern (Djibouti, Mauritius, Somalia, South Sudan), Southern (Botswana) and Western (Guinea Bassau). For those risk factors with multiple variables, the variables with the most complete data were selected. Statistical analysis then determined which variables were included in the final model.

### 2.5. Statistical analysis

Regional IMR data, cervical cancer prevention programmes and coverage data were described through IQR boxplots, with Mood's median test and *post-hoc* Pairwise Median test performed on each. Risk factor data from the 46 countries were initially examined for normalcy and outliers, through Shapiro-Wilks test. Descriptive statistics for each variable can be seen in Table 3. The distribution of the outcome variable (IMR) was assessed, and the skewness detected was corrected through a $Log_{10}$ transformation and confirmed with a QQ-plot. Pearson correlation coefficients for those variables of normal distribution, Spearman correlation coefficients for those non-parametric variables, and probability values ($p$) were generated and used to test the association between each risk factor and the logIMR. All $p$-values presented are two-sided with significant probability set at $\alpha < 0.05$. Co-linearity between the relevant risk factors was assessed through a scatterplot matrix. In the case of co-linearity, a single variable consisting of a large $n$ value, high regression coefficient and a low $p$-value was selected from those co-linear risk factors. Of the remaining risk factors showing a statistically significant correlation with the logIMR, there were a number of missing values (14.2%). To retain as much of the data as possible, the missing values were calculated using mean imputation with additive regression (an additive cubic spline), bootstrapping and predictive mean matching. Through different bootstraps, data were randomly resampled for each of the multiple imputations, followed by fitting one of ten flexible additive models to the samples with replacements from the original data. Finally, these additive models predicted the missing values by using the non-missing values, and the results then pooled. Two categorical variables were then added to the dataset: *Region* ("Southern", Eastern", "Central" and "Western") and *Screening* ("Population-based", "Opportunistic", "No programme" and "No data"), and the complete dataset underwent a stepwise multiple linear regression to select the combination of variables to describe late-stage cervical cancer for SSA. Additionally, a stepwise regression was performed using the non-missing data and results compared with the data imputation model. Regression diagnostics of the final model was conducted using ANOVA Type II sum of squares tests assessing any interaction between variables. All data analyses were performed using R version 3.4.1 (2017-06-07) for Mac iOS with the following packages: *pastecs, MASS, psych, car, gvlma, RVAidememoire, rcompanion, RColorBrewer, rworldmap, rworldxtra* and *Hmisc.*

## 3. Results

### 3.1. Incidence-to-mortality ratio

The geographic distribution of country-specific cervical cancer incidence, mortality and IMRs appear in Fig. 1a–c, with a comparison of regional and individual IMRs in Fig. 2a and b. Overall, the IMR for sub-Saharan Africa is 1.569. By region, Western Africa has the lowest median IMR at 1.463, followed by Eastern Africa (IMR = 1.595) and Central Africa (IMR = 1.675), whereas Southern Africa is highest with 1.761. The effect of cervical cancer prevention programmes on the IMR, as well as the coverage achieved, can be seen in Fig. 3a and b. The