

Chapter 23

Data Science in Global Health—Highlighting the Burdens of Human Papillomavirus and Cervical Cancer in the MENA Region Using Open Source Data and Spatial Analysis



Melek Somai, Sylvia Levy, and Zied Mhirsi

Abstract Cervical cancer is a top driver of death and disability across the MENA region with at least 7,601 deaths annually. Nearly all cases of cervical cancer are caused by Human papillomavirus (HPV), the most common viral infection of the reproductive tract. HPV infection can be prevented by widespread uptake of the HPV vaccine and progression to cervical cancer can be averted with regular HPV and cervical cancer screenings. Sadly, these effective interventions are not in broad use on a national and regional level in the MENA region. We developed a data-driven digital map that integrates multiple data sources about HPV vaccination and cervical cancer incidence and mortality for countries in the MENA region. The use of different data sources from international and national organisations offers integrative and comprehensive information about the epidemiological status of these preventable diseases and the current policy-effectiveness at the national level. Our platform is a one-stop analytical online application that can help policymakers in their decision-making and ease the process required to combine different data sources into a comprehensive platform.

Keywords Human papillomavirus · Cervical cancer · Data mashup · GIS

Learning Objectives

By the end of this chapter, you will be able to:

- Understand the burden of Human Papillomavirus and Cervical Cancer in the MENA Region.
- Enumerate the different challenges in establishing a data-driven approach for global health policy.

M. Somai (✉)

Collaborative for Healthcare Delivery Science, Medical College of Wisconsin, 8701 Watertown Plank Rd., Milwaukee, WI 53226, USA

e-mail: msomai@mcw.edu

S. Levy · Z. Mhirsi

Global Health Strategies, 38 East 32nd Street, 12th Floor, New York, NY 10016, USA

© The Author(s) 2020

L. A. Celi et al. (eds.), *Leveraging Data Science for Global Health*,
https://doi.org/10.1007/978-3-030-47994-7_23

373

- Describe the process to leverage open source data to build an online digital dashboard.
- Evaluate the challenges and the limitations of a data science approach in global health.

23.1 Introduction

Countries across the Middle East and North Africa (MENA) region are facing a set of unique challenges, including the growing burden of Non-Communicable Diseases (NCDs), compounded by disparities in access to affordable and equitable health services (Middle East and North Africa Health Strategy 2013–2018). NCDs and injuries account for more than 75% of total disability-adjusted life years (DALYs).

Cervical cancer is one such NCD and is a top driver of death and disability across the MENA region. Between 2012 and 2018, the number of deaths every year due to cervical cancer doubled in most countries in the MENA region, as defined by UNAIDS. Today, cervical cancer causes at least 7,601 deaths annually in the region (Cancer today 2019). If decisive steps are not taken at the national and regional levels, annual deaths due to this preventable disease will double again by 2040, reaching 15,728 deaths per year across the MENA region (Cancer tomorrow 2019).

Nearly all cases of cervical cancer are caused by Human papillomavirus (HPV), the most common viral infection of the reproductive tract. In particular, HPV types 16 and 18 cause approximately 70% of invasive cervical cancers (Cancer today 2019). HPV also causes infections and cancers in other areas. Across countries in the MENA region, HPV prevalence rates vary. Some studies show that more than 21.1% of women in the general population of some MENA region countries have HPV type 16 or 18 at a given time (Cancer today 2019; Bruni et al. 2018).

HPV infection can be prevented by widespread uptake of the HPV vaccine and progression to cervical cancer can be averted with regular HPV and cervical cancer screenings. Countries must implement a comprehensive approach to these deadly diseases, incorporating preventive measures, as well as community education and awareness efforts, early and high-quality treatment for cervical abnormalities, cervical cancer and other cancers related to HPV infection, and palliative care. Sadly, these effective interventions are not in broad use on a national and regional level in the MENA region.

Annually, at least 11,202 women in the MENA region are newly diagnosed with cervical cancer (Cancer today 2019). Across countries in the MENA region, incidence and mortality rates vary [Fig. 23.1]. Somalia and Morocco have some of the highest incidence and mortality rates, with 24.0 and 17.2 women per 100,000 being newly diagnosed with cervical cancer annually and at least 21.9 and 12.6 women per 100,000 dying due to cervical cancer per year, respectively. Whereas Iran, Iraq and Yemen have the lowest (around 2 per 100,000 women are diagnosed per year and about 1 per 100,000 die because of cervical cancer annually) (Cancer today 2019).

Cervical cancer across the MENA region

Estimates as of 2018 ● Death rate (per 100,000 women) ● Incidence rate (per 100,000 women)

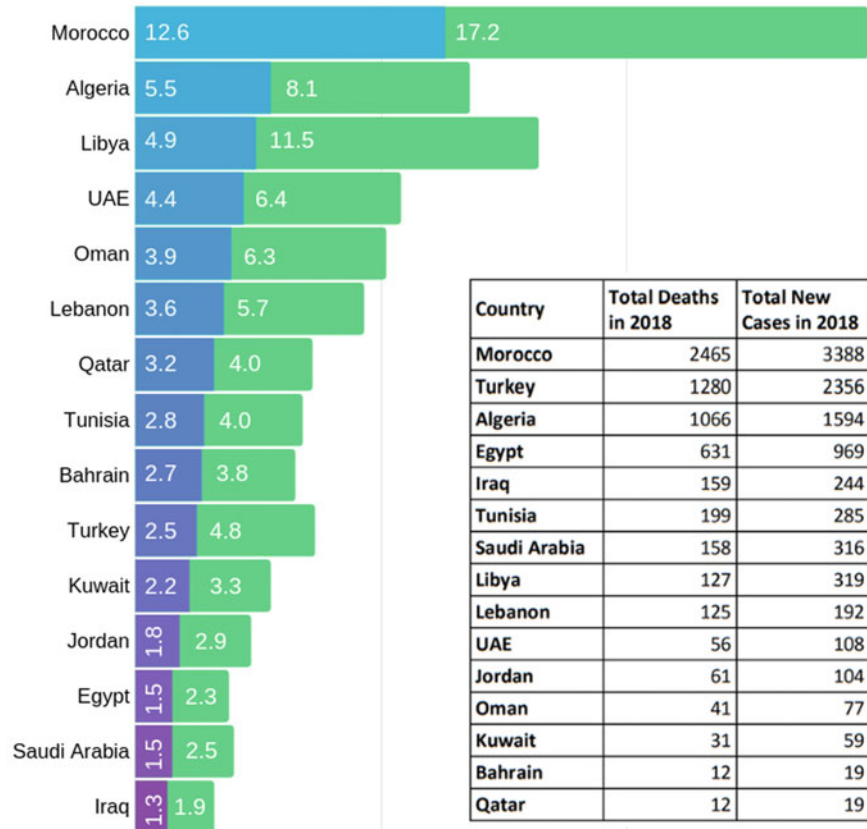


Fig. 23.1 Cervical Cancer Incidence and Death Rate per 100,000 women across the MENA Region in 2018

Cervical cancer incidence and mortality rates are, on average, lower in the MENA region compared to the rest of the globe. Therefore, scaling up the right preventive and care interventions at the national and regional levels could potentially steer the region toward HPV and cervical cancer elimination. But without early and effective action against HPV and cervical cancer, these rates could increase quickly and elimination may become much more difficult.

Despite this disease burden and the current opportunity for elimination, only two countries in the MENA region have integrated the HPV vaccine into their national vaccination programs—the United Arab Emirates (UAE) and Libya. Other countries, including Morocco, have announced their plans to roll out the vaccine in the near future (Internet 2019). Discussions around HPV and cervical cancer prevention in

the region have been ongoing, with a number of countries raising the importance of these preventable diseases on their national health agendas and carrying out local and national cervical cancer screening campaigns. However, the lack of leadership and clear action from the majority of countries in the region risks future increases in annual deaths and new cases of cervical cancer across the MENA region.

While the MENA region as a whole is experiencing HPV and cervical cancer epidemic, a one-size-fits-all solution would be inappropriate, considering distinct political, economic and social contexts across the region. In terms of health spending, MENA region countries spend on average 5.3% of their Gross Domestic Product (GDP) on healthcare, an abysmally low figure in comparison with the global rate of 8.6% of total expenditure on health as a share of gross domestic product. Moreover, the region is characterized by a high share of out-of-pocket expenditure for health services which represented 35% of the entire healthcare spending in 2013 [13–76%], when the average in OECD countries is 13%. Therefore, with these constraints in health financing and the catastrophic burden of diseases, countries in the MENA region require a tailored approach to improving healthcare policy decision-making and the redesign of healthcare services (Asbu et al. 2017).

To ensure public health interventions to stem the tide of HPV and cervical cancer are successful in the region, policies must evolve alongside rigorous monitoring of effectiveness, accessibility and applicability. Any effort to scale up training for health workers, implement new practice in managing preventive service delivery, or launch community-based interventions must be informed by data and evidence from the affected communities. In this regard, it is important for policymakers and researchers to have easy access to key data points from health, demographic and epidemiological surveillance systems (Lang 2011). This data-driven approach offers a “data-first” feedback mechanism that can transition the current public health systems in the MENA region toward evidence-based practice and policy design. In addition, integrating these data sources with spatial analysis offers a revolutionary way to explore public and global health data. Indeed, Geographic Information Systems (GIS) and related information and mapping technologies are considered by a recent WHO report as “the forefront of cutting edge tools that are being used to build reliable public health information and surveillance systems” (Organization WH 2006).

While data sources on a wide range of public health issues, including HPV and cervical cancer, are available online, accessing and combining these data sources is not straightforward. Most data sources use different data structures, terminologies and semantics. Combining these data points is time-consuming and technically challenging (Butler 2006). Moreover, considering the opportunities for data sharing and integration of GIS technology in the region, and low and middle-income countries in general, the lack of development in the field of data science represents a large impediment in the MENA region (Lang 2011).

23.2 Methods

The process of implementation of our dashboard relied on the Agile methodology in combination with a Human-Centered Design approach during the design and ideation phase. The Agile methodology allowed the team to iterate quickly and improve the platform incrementally. A data-driven approach to design the solution is impossible if we had to follow a bottom-up approach. In other words, in order to let the data guide our design decision and to connect the data insights to the challenges of HPV and cervical cancer in the region, we followed an iterative and agile process and tried to limit our own bias when reviewing and curating the data, and designing the dashboard. In the following section, we describe this approach.

23.2.1 Data-Curation

We sourced data on the regional burden of cervical cancer from the International Agency for Research on Cancer (IARC) Global Cancer Observatory (GCO) estimates of incidence, mortality and prevalence for the year 2018 in 185 countries or territories for 36 cancer types by sex and age group. After collecting annual cervical cancer incidence and mortality rates from each country in the MENA region, as defined by UNAIDS, from this database, as well as projections for each of these data points through 2040, we calculated regional totals to use in our data visualisation. We also used country reports put together by the HPV Information Centre and news coverage to determine which countries across the region have implemented the HPV vaccine.

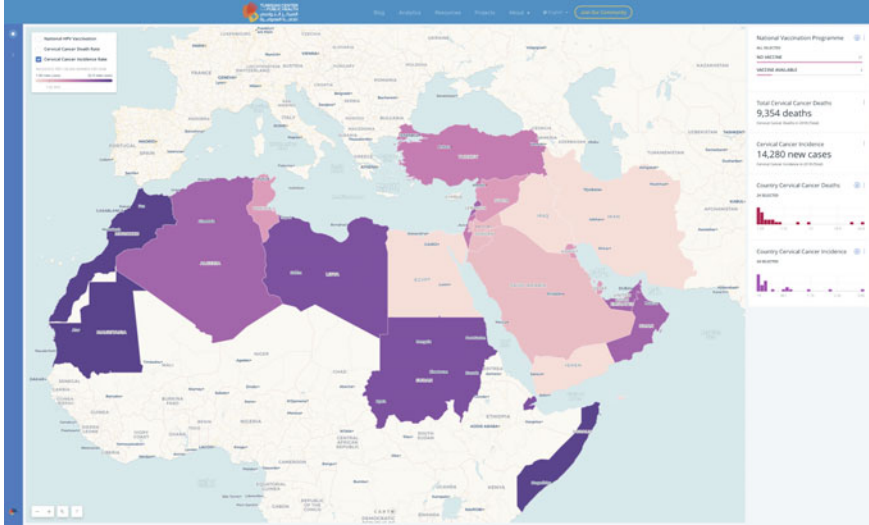
23.2.2 Data-Mashup

In order to combine the different data sources, we used the international country code ISO 3166-1 alpha-3 (ISO-3) developed by the International Standard Organization. In case a data source did not include the ISO-3 code, we used the R-Package ‘*countrycode*’ (version 1.1.0) which converts the country’s name to ISO-3. We relied on the *countrycode* package to reduce human errors and ensure that the process of generating the dataset is automated. This enables reproducibility and replicability of our data pipeline.

23.2.3 GIS Platform

To develop the GIS, we used Carto technology (<https://carto.com/>). Carto is a cloud-based platform used to build powerful “Location Intelligence” applications. It

includes a data repository that converts ISO-3 codes to spatial data and generates a spatial data representation on an interactive map. Carto GIS interface is customisable and includes several features such as multi-layered mapping, dynamic statistics and filtering.



23.2.4 Dashboard and Profile Page Development

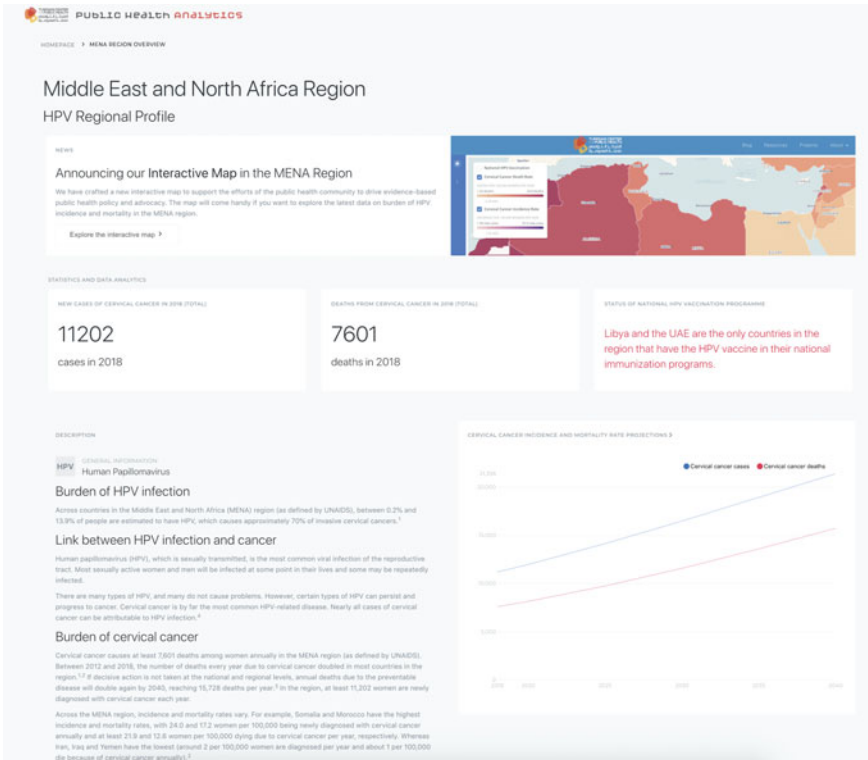
After we developed the interactive map, the second phase of the project was the development of a regional page and country profile web pages, each with three main sections, to provide a narrative and additional data points for viewers. The first section of each of these pages includes a “dashboard” which highlights the current status of several key performance indicators (e.g. new cases of cervical cancer in 2018, deaths from cervical cancer in 2018 (total), the status of the national HPV vaccination programme). These data points came from the IARC GCO estimates and the HPV Information Centre’s database. The second section of each page provides a short description of the current epidemiological challenges facing that country in terms of the burdens of HPV and cervical cancer. The third section lists the peer-reviewed publications and news articles that relate to the context of these preventable diseases in each specific country, touching on current levels of awareness, effective interventions and actions at the local and national levels and future approaches to prevention.

Data points on HPV prevalence, as well as the lists of peer-reviewed publications from across the region, were collected through a thorough internet search of recent academic journal articles discussing “HPV” and/or “cervical cancer” in each country.

Bringing together over 250 academic publications and data sources, the regional and country profile pages serve as a useful repository of information and insight into the current landscape of disease burden, infection and awareness in the MENA region.

We sourced relevant news articles on each regional and country profile page from internet searches for pieces covering “HPV” and/or “cervical cancer” in Arabic, English and French. Given the large quantity of media coverage, we have included the latest and most relevant pieces to each country and the region’s HPV and cervical cancer prevention efforts. An important point to consider is that although the inclusion of news articles could provide a more up-to-date view of current realities, the validation of the content in the news articles may be challenging especially with the rise of false or misleading news. While, we understand these limitations, we identified a process of quasi-peer-review as a workable solution. In this quasi-peer-review process, two independent members of the group validated the content of the articles. In a later stage, we consider including a time-dependent factor and excluding news articles that were not published within the last year. This process is time consuming, and other alternative such as crowd-sourcing could be potentially explored.

The dashboards and profile pages were developed by the engineering team at the Tunisian Center for Public Health and uses a Django framework, which can be used to develop websites and web applications and uses the Python programming language. The data enrichment pipeline is based on R programming language.



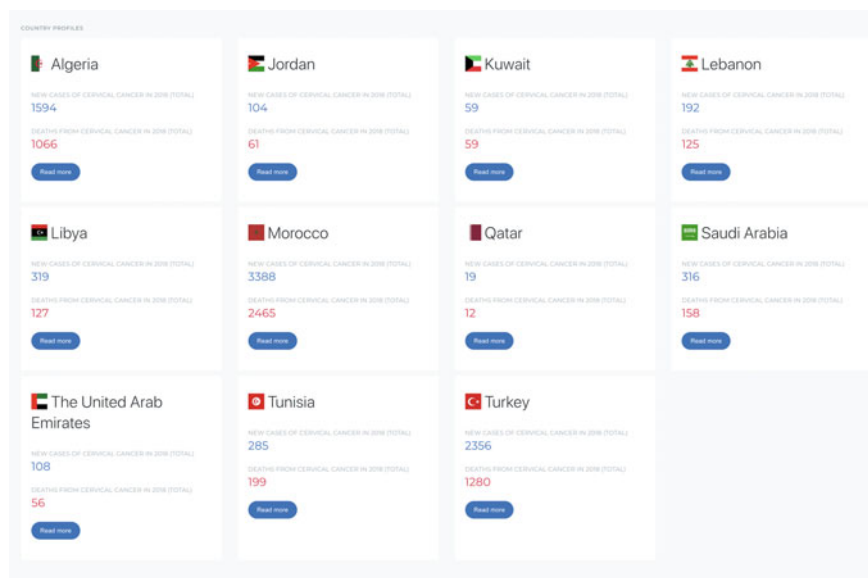
23.3 Findings

23.3.1 *Challenges and Limitations*

Our core challenges in developing the interactive map and regional and country profile pages centered on the lack of national-level data on HPV prevalence and the relative prevalence of each HPV strain. Academic studies on community and city-level burdens of HPV infection were incorporated into our platform, but because there are few national-level data points available, any analysis using the interactive map and regional and country profile pages is limited to what little is known about HPV prevalence. In the next iteration of the map, we may explore integrating data from the Institute for Health Metrics and Evaluation on Disability Adjusted Life Years due to cervical cancer in the MENA region. Similarly, few academic studies on the cost-effectiveness of and potential pathways to the rollout of nationwide HPV vaccination and cervical cancer screening programs exist. Future research in this area is needed to elucidate the context of HPV and cervical cancer prevention at the country level and inform policy. With additional data and analysis on options for the way forward for HPV vaccination and cervical cancer prevention, the map and profile pages could provide a much more complete picture for policymakers and advocates who might use it.

Another key challenge was the fact that the data we highlight on annual and projected country-level cervical cancer incidence and mortality rates come from a repository of estimates, rather than concrete datapoints—the IARC GCO database. On the GCO online platform, the authors highlight that the data points they present “are the best available for each country worldwide. However, caution must be exercised when interpreting the data, recognizing the current limitations in the quality and coverage of cancer data, particularly in low- and middle-income countries.” Our interactive map and profile pages are therefore limited and should be used carefully given the possibility that the IARC GCO estimates of cervical cancer incidence and mortality in countries in the MENA region could be inaccurate to the reality on the ground.

A few more limitations do exist in the current iteration of the interactive map and regional and country profile pages. The process for data validation and quality control is difficult and not totally transparent. Moreover, while the application is online, policymakers and researchers might lack the technical expertise to analyse the data, navigate the GIS system and derive accurate information. An effort to educate, train and support users continuously is important. Therefore, an important feature to include in the upcoming release is the development of an online help center and a repository of training resources. The map and profile pages are also intended to be a platform, rather than a definitive authority on HPV and cervical cancer in the MENA region. Into the future, this platform will grow and our team will continue to collect relevant data points from international and national sources, academic articles and news pieces.



23.4 Conclusions

Through this project, we developed the first data-driven and digital map that integrates multiple data sources about HPV vaccination and cervical cancer incidence and mortality for countries in the MENA region. Our interactive map and regional and country profile pages are a powerful digital platform for policymakers, academics and advocates to utilise. The use of different data sources from international and national organisations offers integrative and comprehensive information about the epidemiological status of these preventable diseases and the current policy-effectiveness at the national level. It also offers a way to compare countries in terms of their policy and disease burden status.

Our platform is a one-stop analytical online application that can help policymakers in their decision-making and ease the process required to combine different data sources into a comprehensive platform. By developing the profile pages and the map at the regional and national levels, this resource is already being used by local governments, not-for-profit and other international organisations to advocate for better management and policy design to eliminate HPV and cervical cancer in the MENA region.

References

- Asbu, E. Z., Masri, M. D., & Kaissi, A. (2017). Health status and health systems financing in the MENA region: roadmap to universal health coverage. *Global Health Research and Policy*, 2, 25.
- Bruni, L., Barrionuevo-Rosas, L., Albero, G., Serrano, B., Mena, M., Gómez, D et al. (2018). ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). Human Papillomavirus and Related Diseases in Americas. Summary Report 10 December 2018.
- Butler, D. (2006). Mashups mix data into global service. *Nature*, 439, 6–7.
- Cancer today. (2019). <https://gco.iarc.fr/today/home>. [cited 15 Dec 2019].
- Cancer tomorrow. (2019). <https://gco.iarc.fr/tomorrow/home>. [cited 15 Dec 2019].
- Lang, T. (2011). Advancing global health research through digital technology and sharing data. *Science*, 331, 714–717.
- Middle East and North Africa Health Strategy. (2013–2018). In: World Bank [Internet]. <https://www.worldbank.org/en/region/mena/publication/mena-health-strategy>. [cited 15 Dec 2019].
- Organization WH. (2006). Others. Public health mapping and GIS for global health security: WHO strategic and operational framework. Geneva: World Health Organization. https://apps.who.int/iris/bitstream/handle/10665/69715/WHO_CDS_GIS_2006.1_eng.pdf.
- مغرس: وزارة الصحة تعزز تلقيح الفتيات ضد سرطان عنق الرحم منذ الطفولة [Internet]. (2019). <https://www.maghress.com/bayanealyaoume/126728>. [cited 15 Dec 2019].

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

