|  |
| --- |
| **DATA ANALYTICS AND VISUALIZATION PROJECT PROPOSAL** |
| **Name:** Amy Reidy - s00215964 |
| **Date:** 29/11/2021 |
| **Title:** “An Exploration of Global Inequalities in Cervical Cancer” |
| **Background and Rationale:**  Cervical cancer is the fourth most common cancer affecting women, and according to the latest statistics from the International Agency for Research on Cancer (2021), there were 604,237 new cases and 341,831 deaths due to cervical cancer in 2020. The number of incidences and the mortality rates of cervical cancer have been slowly decreasing in the past few decades. However, in some countries (such as Lesotho, Zimbabwe, and Bulgaria) the number of new cases is actually increasing, and women in low-resource settings are still disproportionately affected by cervical cancer (Zhang et al., 2021).  Despite cervical cancer being so common, it is an extremely preventable type of cancer. 99% of cervical cancers are caused by the high-risk human papillomaviruses (HPV), and immunization against HPV has been proven to be highly effective at preventing the development of cervical cancer (Future II Study Group, 2007). Screening for and treating precancerous lesions is also an extremely effective secondary prevention approach, and the earlier that the cancer is diagnosed the easier it is to treat and the higher the survival rate.  However, low-income and middle-income countries (LMICs) tend to have less access to these preventative services compared to richer countries, as well as less access to treatment for the later stages of the disease (Anaman-Torgbor et al., 2020). As a result, there are much higher mortality rates in these countries, and of the 310,00 women who died of this disease in 2018, 90% were living in LMICs (Fray et al, 2020). |
| **Objective and Research Questions:**  The goal of this project is to increase awareness of global inequalities in cervical cancer incidence and mortality rates by exploring and visualizing data related to this disease.  Through both the literature review and the analysis of relevant datasets, this project aims to answer these research questions:   * How do the most recent rates of cervical cancer vary across the world? * Which countries/regions have the highest rates of cervical cancer (cases and mortality)? * What causes cervical cancer and why is it more prevalent in certain countries? * Why do certain countries have higher mortality rates for cervical cancer? * How do cervical cancer rates compare to rates of other cancers across the world? * How do HPV vaccination rates vary across the world? * Is there a relationship between HPV vaccination rates and cervical cancer incidence rates? * Is there a relationship between poverty rates and cervical cancer mortality rates?   The answers to these questions will be presented in a literature review and in the form of an interactive dashboard that allows users to filter the information. |
| **Methodology:**  This project will use both the literature review and analysis of public datasets to answer the research questions.  Datasets have been obtained from www.gco.iarc.fr/projects, www.hdr.undp.org and www.path.org  Please note that while this project will refer to data about women, cervical cancer can affect any person who has a cervix. Unfortunately, there is not global cervical cancer data available for trans-men, non-binary or other gender expressions, and it is unclear if they have been included in the statistics for women.  The exploratory data analysis and visualization will be done using Python programming in a Jupyter notebook. An interactive dashboard will be created using Plotly and Dash. |
| **Plan of Work:**   * Preliminary research on topic * Obtain appropriate datasets * Deeper research for literature review   + **Output 1: Final draft of literature review** * Data Preprocessing * Create visualizations of global data related to cervical cancer   + **Output 2: Individual visualizations that display and compare cervical cancer rates for different countries/regions** * Create a dashboard that allows users to view multiple visualizations at once and filter the data in the visualizations (Possible filters: by country, region, year, etc.)   + **Output 3: Project dashboard(s)** |
| **References:**  Anaman-Torgbor, J., Angmorterh, S. K., Dordunoo, D., & Ofori, E. K. (2020). [Cervical cancer screening behaviours and challenges: a sub-Saharan Africa perspective.](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7392861/) The Pan African Medical Journal, 36.  Future II Study Group (2007). [Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions.](https://www.nejm.org/doi/full/10.1056/nejmoa061741) *New England Journal of Medicine*, *356*(19), 1915-1927.  International Agency for Research on Cancer (2021). [Cervix uteri (Fact Sheet)](https://gco.iarc.fr/today/data/factsheets/cancers/23-Cervix-uteri-fact-sheet.pdf).  Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). [Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.](https://acsjournals.onlinelibrary.wiley.com/doi/full/10.3322/caac.21492) CA: a cancer journal for clinicians, 68(6), 394-424.  Zhang, X., Zeng, Q., Cai, W., & Ruan, W. (2021). [Trends of cervical cancer at global, regional, and national level: data from the Global Burden of Disease study 2019.](https://bmcpublichealth.biomedcentral.com/track/pdf/10.1186/s12889-021-10907-5.pdf) *BMC public health*, *21*(1), 1-10. |