STEM Fellowship Big Data Challenge 2020

Effectiveness of Diagnostic and Treatment Modalities for Breast and Prostate Cancer in Countries with Variable Levels of Development

Fedrigo, R[1, 3], Haid, E[2], Watkins, A[2, 3], and Winata, H[1, 3]

[1]University of British Columbia, Vancouver, BC, Canada
[2]Simon Fraser University, Vancouver, BC, Canada
[3]BC Cancer Research Institute, Vancouver, Canada

June 1, 2020

Abstract

Chronic health conditions, such as cancer, have applied significant strain on health care systems in recent years, likely due to early detection and improved disease prognosis. Breast and prostate cancer both represent a considerable portion of chronic health conditions, as they are characterized by high incidence in the population and long survival. Despite the existence of successful methods for screening and diagnosing breast and prostate cancer, few studies have investigated the global impact of screening tools on patient survival rates and their economic repercussions. This paper aims to investigate how the implementation of diagnostic, therapeutic, and screening technology, for both breast and prostate cancer, impacts countries based on their level of development. Datasets were used to classify 194 nations based on their human development index (HDI). Quantities of diagnostic (MRI, CT, PET, gamma cameras) and therapeutic (linear accelerators and cobalt machines) modalities were correlated with relative mortality rates in each country, using the Pearson correlation. Strong, negative correlations were found between the number of machines and relative mortality rates, particularly for countries with an HDI in the top 25%. Health care spending had a stronger correlation with the number of therapeutic, rather than diagnostic, machines (R^2=0.59 and R^2=0.20 respectively). Overall, developed countries with high investments in diagnostic machinery, did not spend more in health care. Since it is well-known that early detection of cancer is associated with improved survival and quality of life, it can be asserted that investment in diagnostic machinery is beneficial from both a patient care and economic perspective. The widespread availability and success of national screening protocols was highlighted for breast cancer; indicating a possible unmet need for prostate cancer.

Keywords
cancer, early detection, public health

1 Introduction

As the life expectancy of the world population increases, complex and chronic diseases such as diabetes, heart disease, mental illness, autoimmune disorders and cancer have become more prevalent. The demographics of certain countries such as Canada, Japan and Germany, have shifted towards older populations, resulting in cancer becoming an ever-increasing burden to both the country's citizens and its healthcare system. In 2018, the cancer mortality rate was estimated to be approximately 9.6 million globally, with approximately 70% of deaths occurring in low-to-middle income countries [1].

In addition to the high mortality rate associated with cancer, cancer also applies a significant strain to both health care systems and the economy. In 2015, the direct cost of treating cancer in the United States, a country which devotes significant resources towards cancer treatment, was estimated to be $80.2 billion [2]. In contrast, developing countries are historically adapted towards treating infectious and parasitic diseases rather than chronic conditions such as cancer [3]. This results in developing countries being ill-equipped to treat cancer, a disease that often requires extensive intervention and long-term treatment plans [3].

While evaluating the effects of cancer, it is critical to acknowledge that the term is used to broadly refer to over 200 distinct and heterogeneous diseases [4]. Some types of cancer are associated with poor prognosis and low rates of survival, while others are slow-growing and easily treatable if detected early. For instance, breast and prostate cancers have relatively high 5-year survival rates (89.7% and 98.6 % respectively) [4].

Most developed countries have adopted various methods of prevention and screening, providing early diagnosis of cancer. Prevention focuses on limiting risk factors such as smoking, radiation, and carcinogens. For example, many governments have implemented vaccinations to limit the spread of human pappillomavirus (HPV), a virus directly linked to cervical and oropharyngeal cancer [5]. Meanwhile, screening and early diagnosis are part of an early detection strategy. Early detection has been shown to have increased the 5-year cancer survival rate in the U.S. between 1970-77 and 2007-13 [4]. Screening procedures often involve applying simple tests on general population so that diseases can be identified before symptoms manifest. At the very least, screening procedures aim to diagnose individuals with cancer at an earlier stage of disease progression, improving the likelihood that the cancer is treatable. Overall, screening measures involve both educating the public to recognize early cancer symptoms, and providing the correct diagnostic procedures and technology that allows the confirmation cases.

Breast and prostate cancer is the most common cancer diagnosis for females and males respectively, both contributing significantly to cancer-related deaths of each sex, following only lung and colorectal cancer [4] [6] [7]. Breast cancer can be screened using mammograms, while prostate cancer is detected through digital rectal exams (DRE) or by measuring prostate-specific antigen (PSA) in blood. Survival rates of breast and prostate cancer depend significantly on the timing of diagnosis as response to treatment is inversely correlated with disease progression [7].

Despite the technological ability to perform early screening, the prognosis associated with breast and prostate cancer varies considerably between countries. These countries differ in income level and public health policies. It is the aim of this paper to investigate, on a global scale, how the availability of public health screening procedures is related to the incidence and mortality rate of both breast and prostate cancer in the general population. Furthermore, this paper aims to investigate the economic benefit and or burden associated with implementing early screening procedures in both developing and developed countries.

2 Materials & Methods

Breast and prostate cancer were selected due to their frequency in the population, as well as the fact that they primarily impact both female and male populations independently [4]. Furthermore, breast and prostate cancer have widely available screening methods, treatment methods, and the survival rates are highly correlated with early detection [6] [7].

We used datasets obtained from Gapminder and the Global Health Observatory (GHO) database, provided by the World Health Organization (WHO). Data sets were selected to provide insight on the level of development, the availability of cancer screening, the availability of early diagnosis measures, as well as other cancer statistics in each country.

The Human Development Index (HDI) [8], used in this paper as a proxy measure for social development and economic growth, is based on the Gross Domestic Product (GDP) and Gross National Income (GNI) per capita, and other socio-economic factors [8].

Availability of screening and early diagnosis procedures in each country are evaluated based on the number of diagnostic and treatment devices commonly used for breast, prostate and other types of cancer [9]. Computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and gamma cameras are commonly used to detect and quantify tumours in cancer patients. Meanwhile, linear accelerators (LINACs) and cobalt machines are used in external beam radiation therapy treatment. Additionally, the existence of national screening programs for breast cancer [10] are used to indicate the availability and accessibility of breast cancer screening. Data for the variables are obtained for as many countries as possible from the World Health Organization database.

To aggregate all relevant information from a multitude of diffident databases, a significant amount of prepossessing had to be done. All data processing was done using Python (version 3.8) and Pandas (version 1.0.3), a Python data analysis library. Raw data was collected in the form of both CSV files and scraped JSON web requests. Python code was written to parse JSON web requests into CSV files structured similarly to the other datasets.

Once all the raw data was in a consistent CSV file format, two main aggregated data sets were created. The first data set was indexed by country and year, while the second was indexed solely by country. The dataset indexed by country and year contained data points taken from the year 2000 up until the most recently reported year, often 2019, for each country. The data set was then collapsed into a dataset indexed only by country in order to be joinable with other aggregated dataset indexed only by country. When collapsing, the value most recently reported by each country was taken for each field, most often falling between 2014-2016. This was a necessary step as some fields are not reported annually. Next, fields with values such as 'N/A', 'Not reported', and 'unsure' were all normalized to null for consistency. Prior to analysis and plotting, countries (rows) with a missing HDI index value or more than one null value were also removed from the data set. The final data set contains information from 194 countries (rows) and 34 variables (columns) alongside with information from 2000 - 2019 in the year-indexed data set.

After the aggregation and formatting of raw data was complete, additional columns were created to extend this paper's investigation. The countries were categorized into quartiles based on their HDI value where the 1^{st}, 2^{nd}, $3rd$ and 4^{th} quartile are analougous to the $0-24^{th}$, $25-49^{th}$, $50-74^{th}$ and $75-100^{th}$ percentile. The quartiles corresponds to HDI values of 0-0.56, 0.56-0.73, 0.73-0.804 and 0.804-1 respectively. This designation was added as a field in the data set. Additionally, a Python dictionary of continents and countries was used to create a column linking each country to its own continent. Lastly, the relative mortality rate of both breast and prostate cancer were calculated and added as a column by taking the ratio of deaths to new cases.

3 Results

An interactive version of this paper was created and can be found at https://bdc.report. The interactive version allows for dynamic visualization of this paper's data. Plots can be generated based on individual countries, and desired variables can be selected.

Figure 1 visualizes the distribution of HDI quartiles globally. Developed countries are shown to be localized in North and South America, parts of Europe and Oceania, while developing countries are localized in Africa and South Asia. Not all countries are represented in this choropleth due to the lack of HDI data.

As shown in Table 1, the difference in the average number of deaths between each group of development can be primarily attributed to error (breast cancer range = 16.81±7.18 - 19.28±5.99). Therefore, the difference in relative mortality is mainly due to differences in the number of newly-diagnosed cases.

As shown in Figure 2, strong negative correlations were found between the relative mortality rate and the human development index. R^2=0.83 and 0.86 for breast and prostate cancer respectively. Notably, Myanmar and Kiribati were observed to be significant outliers in the breast cancer relative mortality plot, visualized as blue and green respectively.

Figure 3 shows relative mortality versus the number of medical devices, with data points representing a single country. The plots show a distinct quartile-based clustering; countries with higher development levels (purple) have a lower relative mortality while having a greater number of medical devices per million population.

Percentage of GDP spent on health expenditures for each country is plotted in Figure 4 versus the number of diagnostic and treatment technologies (left and right respectively). All development index quartiles allotted a similar percentage of their total GDP towards health care. As expected, the more developed countries had access to greater numbers of diagnostic and therapeutic machinery. Linear regression models were fitted to data points from each developmental quartile cluster. The most significant correlation occurs between expenditures and number of treatment machinery, in which the highest development cluster (quartile 4) featured a strong-positive trend (R^2=0.59).

The bar plot in Figure 5 visualizes the number of relative breast cancer deaths in different countries from 2014-2016. Countries were categorized based on the presence of a national breast cancer screening program, such as routine mammography. The mean mortality rate of breast cancer in countries with and without screening programs were $44.73 \pm 23.87\%$ and $72.06 \pm 28.92\%$ respectively. The calculated t- and p-value was 4.95 and 1.055e-05 respectively, which indicates a significant difference between the means of the two groups. There were no datasets available for the existence of prostate cancer screening program in each countries from the WHO database.

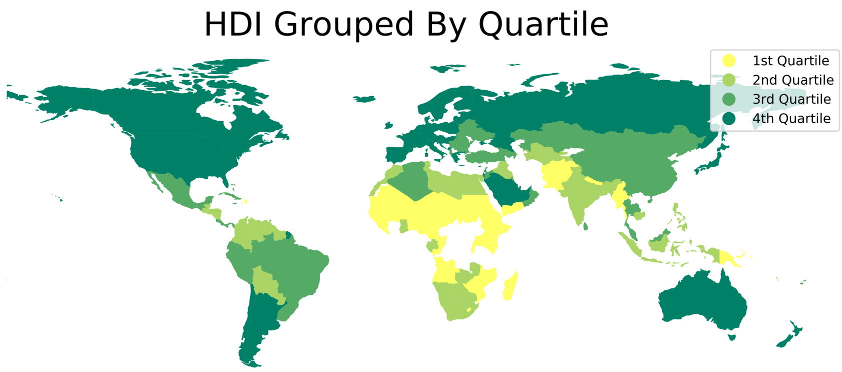

Figure 1: Choropleth map showing the global distribution of country development level as indicated by their HDI quartiles.

Table 1: The mean for breast and prostate cancer new cases, mortality and relative mortality are reported for each development quartile.

Quartile	New cases	Mortality	Relative mortality (%)
Breast cancer per 100k women			
1st	21.83 ±8.39	19.28 ±5.99	91.40 ±14.53
2nd	31.08 ±13.13	16.81 ±7.18	57.07 ±18.40
3rd	51.11 ±21.88	18.38 ±7.49	36.77 ±5.60
4th	80.20 ±21.94	19.13 ±3.46	25.05 ±5.97
Prostate cancer per 100k men			
1st	28.54 ±15.28	32.30 ±16.91	114.1 ±19.79
2nd	39.04 ±42.13	25.85 ±22.37	73.70 ±21.5
3rd	63.70 ±53.95	27.67 ±23.94	45.19 ±9.14
4th	97.26 ±42.76	21.86 ±6.73	25.92 ±10.50

Table 2: R^2 values correlating breast and prostate cancer mortality rates and respective imaging or treatment modality, calculated for countries in each quartile. Bold values indicates $R^2 > 0.1$

	Breast cancer				Prostate cancer			
Modality	1st	2nd	3rd	4th	1st	2nd	3rd	4th
CT	0.075	**0.17**	0.045	**0.25**	0.017	**0.17**	0.015	**0.16**
MRI	0.079	**0.12**	0.016	**0.38**	0.028	**0.13**	0.065	**0.16**
PET	**0.14**	0.077	**0.19**	**0.14**	**0.32**	0.094	0.00052	**0.13**
Gamma	0.029	**0.14**	**0.11**	**0.31**	0.0028	**0.13**	0.065	**0.16**
Diagnostic	0.061	**0.21**	0.038	**0.4**	0.03	**0.19**	0.00032	**0.27**
LINAC	0.025	0.075	0.05	**0.45**	0.012	0.082	0.014	**0.37**
Cobalt	0.074	**0.14**	0.059	0.038	0.048	**0.16**	0.07	0.0024
Treatment	0.086	**0.14**	0.097	**0.44**	0.043	**0.16**	0.076	**0.42**

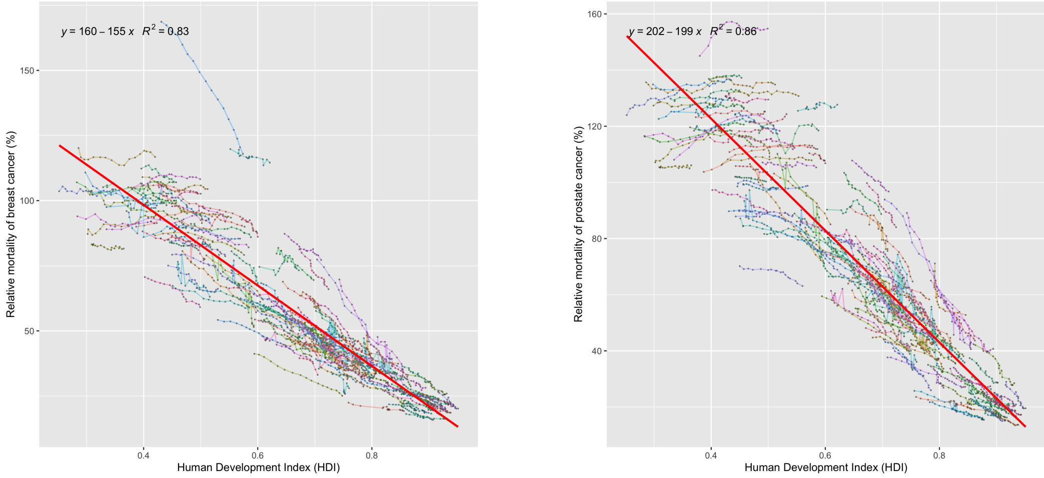

Figure 2: Relative mortality rates (%) of breast and prostate cancer plotted versus human development index (HDI) value. Each colour represents a country over a period of time between 2000 - 2020. Points are fitted linearly, resulting in fit equations of: $RM = 160 - 155(HDI)$ and $RM = 202 - 199(HDI)$, with R^2 values of 0.83 and 0.86 for breast and prostate cancer respectively.

4 Discussion

HDI and Relative Mortality

A strong negative correlation was observed between HDI and relative mortality rates for breast and prostate cancer (Figure 2). As shown in Table 1, most mortality values were within one standard deviation. However, the number of newly-diagnosed cases were significantly higher in countries with higher HDI. It is possible that this trend reflects an increased susceptibility of breast and prostate cancer in the population. As shown by the localization of countries with low and high HDI (Figure 1), environmental factors caused by geographical locations could affect cancer susceptibility. Alternatively, the low numbers of new cases in developing countries could reflect an alarming frequency of undiagnosed cancer cases.

As shown in Figure 2, the relative mortality rate of breast and prostate cancer can be predicted at a population level based on the human development index. The strong negative correlation validates that breast and prostate cancer are highly treatable in countries with higher development. Since relative mortality was computed by dividing the number of new cases by the deaths in a given year, it is possible to attain values greater than 100%. It could also indicate a possible shift in breast and prostate cancer epidemiology, in which the duration of progression-free survival has increased over recent years. These circumstances would result in an increased need for long-term chronic care units for patients diagnosed with breast or prostate cancer [3]. Although health care systems in developed countries have adapted to cope with the health and economic burden of chronic diseases, developing countries do not have the same resources and infrastructure available [3]. Health care systems in developing countries tend to focus on acute treatment that manages infectious and parasitic diseases, with less priority on long-term care [3]. Together, the differing health care strategies between high and low HDI countries could contribute to the pattern seen in Figure 2.

Diagnosis vs Treatment Paradigm

The accessibility and impact of diagnostic technology was assessed by correlating the number of medical devices, such as PET, CT, MRI, and gamma cameras with the relative mortality rate of breast and prostate cancer (Figure 3). As shown in Table 2, weaker linear correlations existed for countries in the 1^{st}, 2^{nd}, and 3^{rd} quartiles of development. Diagnostic devices require a highly specialized team of medical personnel, often comprising imaging technologists, Radiation Safety Officers (RSOs), radiochemists, and medical physicists. Even if sophisticated imaging devices are available, the scanners are not useful without frequent calibration to account for signal drift. Furthermore, scanner images require careful interpretation by a trained radiologist in order to detect tumours. Therefore, the weak correlation between diagnostic devices and mortality rate for low HDI countries, could be explained by the absence of trained medi-

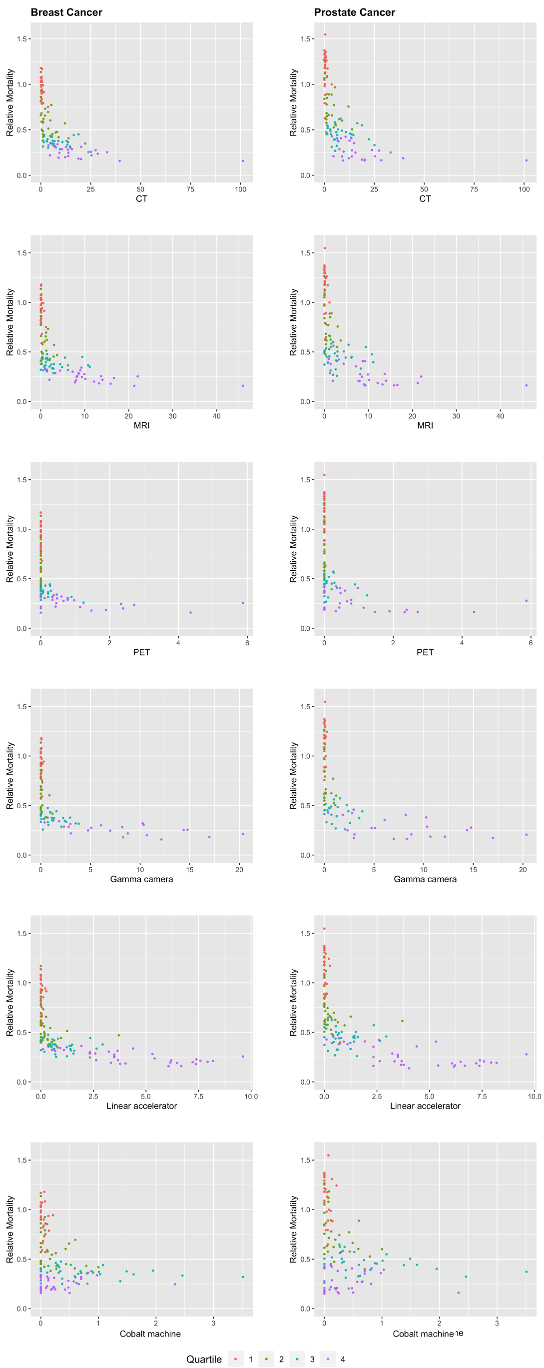

Figure 3: Relative mortality of breast and prostate cancer plotted against number of diagnostic devices (CT, MRI, PET, Gamma camera) and treatment technologies (linear accelerator and cobalt machine) per million population. Each point represents a country, colour-coded by the corresponding development quartile

cal personnel required to maintain and operate these devices.

The weak correlations can also be attributed to limited data provided by countries with decentralized or less thorough cancer registry systems. For instance, medical devices in the United States were absent from this dataset, and supporting literature reviews provided contradictory responses. It is possible that the combined use of public and private health care providers, such as in the United States, resulted in data reporting discrepancies.

In developed countries (3^{rd} and 4^{th} quartile), linear accelerators were most highly correlated with relative rate of mortality in breast and prostate cancer. Meanwhile, in developing countries (1^{st} and 2^{nd} quartile), a stronger correlation was noted for cobalt machines than linear accelerators (i.e. R^2=0.14 and 0.075 respectively). Linear accelerators are considered to be the standard-of-care for external beam radiotherapy in most developed countries, as they can deliver localized and collimated beams directly to the target tissue. Meanwhile, cobalt machines consist of a radioisotope source without an 'off' switch, and have been replaced by linear accelerators in many cases. However, one advantage of cobalt machines is the overall simplicity of the system, which requires less trained employees for operation and maintenance. Therefore, reduced mortality may be more strongly associated with cobalt machines (rather than LINACs) in developing countries with limited resources, since they are more reliable and are easier to operate.

Comparatively, the presence of more diagnostic modalities in highly developed countries was associated with lower rates of relative mortality for both breast and prostate cancer. This trend could reflect the presence of sophisticated imaging departments in highly developed countries. It is critical to note that the only device with a weak trend for developed countries was the cobalt machine, which have been largely phased-out in recent years.

To assess the economical aspect, the annual health expenditure per GDP (%) was used in Figure 4 to approximate a country's investment towards the health sector. As cancer care contributes to the high cost of the health care system [11], it is assumed that the number of devices would be linearly correlated with the percentage of health expenditure per GDP. In the least developed countries (1^{st} quartile), correlation between health expenditure and number of medical devices per million is non-existent. This suggests that regardless of their investment in the health care sector, they do not have the training or infrastructure required to maintain expensive medical devices required for cancer care. Meanwhile, the highest HDI countries (4^{th} quartile) have the strongest correlation between health expenditure per GDP and medical devices per million, with higher significance for investing in treatment technology (R^2 = 0.59).

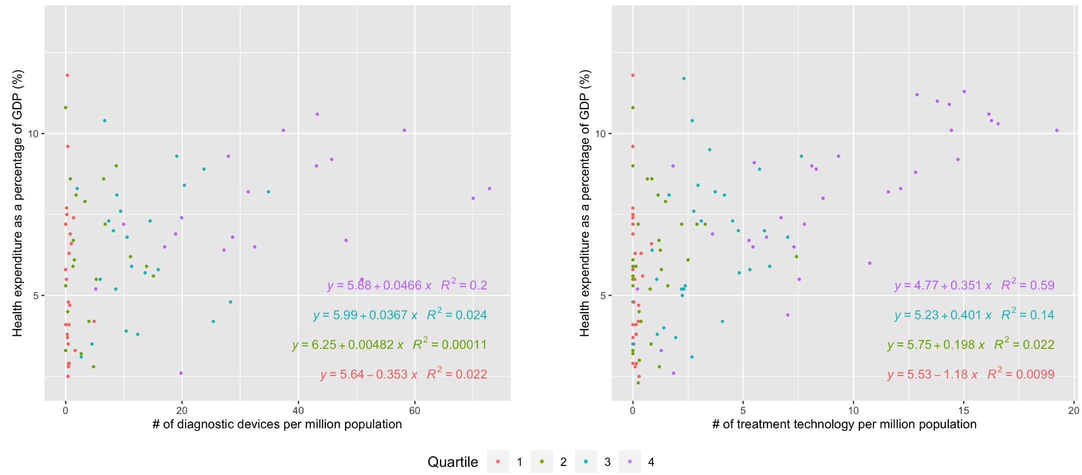

Figure 4: National health expenditure as a percentage of annual GDP plotted against the number of imaging devices **(left)** and treatment technologies **(right)**, for each country. Points are colour-coded based on development quartile, and linear models were fitted for each group. Linear fits and R^2 values shown bottom-right.

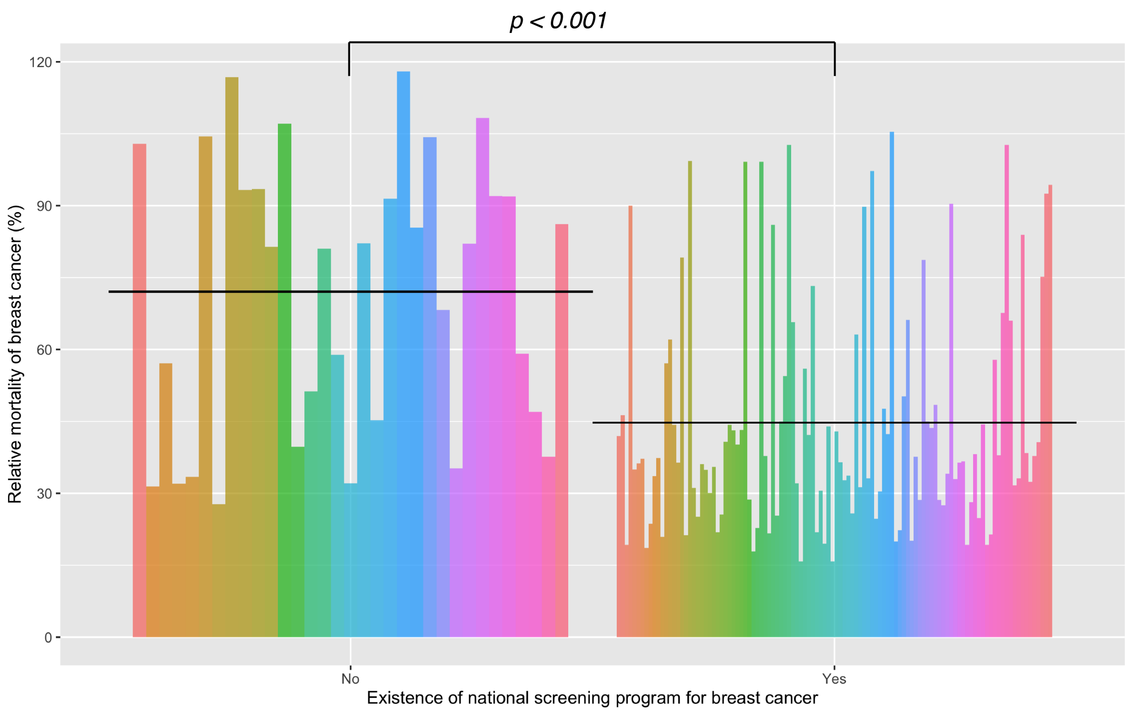

Figure 5: Relative mortality of breast cancer (%) plotted for countries with and without a national breast cancer screening program. Each bar represents an individual country, with 33 and 110 countries with and without a screening program respectively. Horizontal lines represent the mean, $72.06 \pm 28.92\%$ and $44.73 \pm 23.87\%$ for the 'No' and 'Yes' group respectively. *t-test* was calculated to be 4.95 with a *p-value* of 1.055e-05.

Overall, developed countries have invested more in devices intended for diagnosis rather than treatment. From the slopes and R^2 values calculated in Figure 4, we can infer that investing in treatment technology results in greater health expenditure compared with diagnostic devices. This is interesting, given that the infrastructure and regular maintenance required for diagnostic imaging devices, such as PET and MRI machines, are much more intensive than those for linear accelerators and cobalt machines. Therefore, it is possible that early diagnosis - accomplished in part using diagnostic devices, are able to cover their costs by reducing downstream treatment costs. For instance, Stage I breast or prostate cancer are typically treatable with relatively minor surgery or radiation therapy [6]. Conversely, Stage IV breast cancers are less responsive to treatment, and patients are likely to spend more on routine tests and medication required to manage symptoms as the disease progresses [12] [11].

From the bar plots shown in Figure 5, the existence of national screening programs for breast cancer was shown to significantly reduce the mortality rates. Unfortunately, similar data on prostate cancer screening programs are not available from WHO. This could indicate less awareness and education surrounding prostate cancer, relative to breast cancer. Since a strong correlation between relative mortality and implementation of screening programs for breast cancer was observed, there is a significant probability that introducing national screening programs for prostate cancer could result in similar trends. Encouraging healthy men to periodically get a DRE or PSA test could improve the rate of early detection and further reduce the mortality rate.

The fact that there are more countries with a national screening program available for breast cancer reflects the direction taken by health authorities worldwide. Many countries have implemented prevention strategies to mitigate the health, and potentially economic, burdens of breast cancer. Additionally, current research collaborations are in development to use genetic testing to identify new cancer biomarkers. As the costs of next generation sequencing becomes more affordable, screening programs involving genetic testing will become the primary approach. These methods can provide healthy individuals with a better understanding of their risk of cancer, which allows them to make informed medical and lifestyle decisions. Currently, these advancements in screening and early detection procedures would provide the most benefit to developed countries with the infrastructure and resources necessary to treat patients at an early stage. In developing countries, where health care resources are less abundant, our results indicate that a greater number of treatment devices has a more significant correlation with reduced mortality rates in breast and prostate cancer. Therefore, strengthening the health care system to provide the infrastructure necessary for effective treatment would be a better strategy for reducing relative mortality.

Conclusions

The relative mortality rates of breast and prostate cancer were investigated in 194 countries. Strong negative correlations were found between relative mortality and human development index. As a result, countries were grouped into quartiles based on their human development index, and the number of diagnostic and therapeutic devices were compared for each HDI group. The presence of diagnostic and therapeutic devices were both associated with reduced mortality, although the strongest correlations existed for countries with high HDI values. Weak correlations for low HDI countries indicate that simply attaining diagnostic and therapeutic devices is not sufficient. Investment in consistent scanner maintenance and rigorous staff training, as implemented in high HDI countries, is likely necessary to obtain notable improvements in cancer early detection. However, cobalt machines were correlated with reduced mortality in low HDI countries, which may be due to their reliability and practicality. The significant reduction in mortality for countries with national breast cancer screening methods highlights the importance of early detection. However, it also motivates the need to develop parallel national screening programs for prostate cancer.

Acknowledgements

We acknowledge the opportunity to work on this project while competing in the 2020 Big Data Challenge.

Much appreciation goes to Tesla La Touche for her support and guidance throughout this project. Special thanks to Dr. Aaron Goldman for his valuable insight regarding precision medicine. Additional thanks goes to Dr. Carlos Uribe for his input regarding nuclear medicine, and the importance of quantitative PET on predicting patient outcome.

References

[1] Shailja C. Shah, Violet Kayamba, Richard M. Peek, and Douglas Heimburger. Cancer control in low- and middle-income countries: Is it time to consider screening? *Journal of Global Oncology*, (5):1–8, 2019. PMID: 30908147.

[2] American Cancer Society. Economic impact of cancer, Jan 2018.

[3] Alexandra Cameron, Ilse Roubos, Margaret Ewen, Aukje Mantel-Teeuwisse, Hubertus Leufkens, and Richard Laing. Differences in the availability of medicines for chronic and acute conditions in the public and private sectors of developing countries. *Bulletin of*

the World Health Organization, 89:412–21, 06 2011.

[4] Max Roser and Hannah Ritchie. Cancer. *Our World in Data*, 2020. https://ourworldindata.org/cancer.

[5] World Health Organization. Cancer fact sheet, Sep 2018.

[6] Prashanth Rawla. Epidemiology of prostate cancer. *World Journal of Oncology*, 10(2):63–89, Apr 2019.

[7] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6):394–424, 2018.

[8] gapminder database.

[9] Baseline country survey on medical devices, 2014.

[10] Gho | by category | primary and secondary prevention of cancer - response by country, Feb 2020.

[11] Tony Blakely, June Atkinson, Giorgi Kvizhinadze, Nick Wilson, Anna Davies, and Philip Clarke. Patterns of cancer care costs in a country with detailed individual data. *Medical Care*, page 1, 2015.

[12] K Mcpherson. Abc of breast diseases: Breast cancer—epidemiology, risk factors, and genetics. *Bmj*, 321(7261):624–628, Sep 2000.