BIOLOGY HIGHER LEVEL INTERNAL ASSESSMENT
AN INVESTIGATION INTO THE CORRELATION BETWEEN HDI AND MORTALITY RATE DUE TO CORONARY HEART DISEASE IN DEVELOPING AND DEVELOPED COUNTRIES IN EUROPE

RESEARCH QUESTION
What is the correlation between HDI and mortality rate due to coronary heart disease (CHD) per 100,000 people and how does this correlation compare between developing (HDI > 0.788) and developed (HDI < 0.788) countries in Europe?

BACKGROUND INFORMATION
As an aspiring medic I’m always keen to gain a better understanding of the medical profession. I was presented with one such occasion last summer; the opportunity to shadow a cardiologist at the local university’s medical centre for a day. While my day of shadowing was filled with ward rounds and consultations, I was particularly struck by the demographic who presented with diagnoses of coronary heart disease (CHD). Even though the hospital was in the Netherlands, the vast majority of patients with diagnoses of CHD were of Eastern European origin who had recently emigrated to the Netherlands. This got me thinking: “how does one’s origin affect their likelihood of being diagnosed with and succumbing from a certain disease, and what implications might this have for the healthcare industry?”. Ultimately, my shadowing experience inspired me to investigate an epidemiological aspect of medicine in this investigation; the correlation between the Human Development Index (HDI), an indicator of socioeconomic development, and mortality rates due to CHD in developing and developed countries in Europe. Investigating this topic would not only allow me to explore the physiological development of CHD, but also give me a better understanding of the influence of socioeconomic development on disease mortality.

Coronary heart disease is a cardiovascular disease characterized by the occlusion of coronary arteries (the arteries that supply blood to the heart) . This often results in myocardial tissue not receiving sufficient oxygen and nutrients to function, consequently causing a myocardial infarction (heart attack) and an “abrupt loss of blood flow in the body”, known as cardiac arrest . Ultimately, the lack of oxygen supply to the heart as well as other vital organs (e.g. the brain) may lead to irreversible damage, illustrating how CHD may result in death. Critically, CHD is the leading cause of death globally, with an estimated 7.2 million people dying from the disease annually. The main mechanism that results in the occlusion of the coronary arteries is atherosclerosis (the “buildup of cholesterol, fibrous elements, and inflammatory molecules in the walls of arteries”), often restricting blood flow to tissue. This accumulation of cellular substances is often called plaque. An illustration of the impact of plaque on blood flow is shown in Figure 1.

The process of plaque formation in arteries begins with the uptake of low-density lipoprotein (LDL) cholesterol into the subendothelial space in blood vessels. These LDL cholesterol molecules can become oxidized and subsequently engulfed by macrophages, forming so-called “macrophage foam cells” which localize on blood vessel walls and secrete a variety of substances that are responsible for the formation of an atherosclerotic lesion, or “fatty streak”, in a blood vessel. An illustration depicting the process of plaque formation is seen in Figure 2.

These fatty streaks can develop into more complex lesions which are large enough to obstruct blood flow, but the most serious complications arise when these lesions rupture due to, among other reasons, “increased arterial pressure caused by the restricted blood flow and the weakened composition of the lesion”. Ultimately, the rupturing of these lesions exposes the subendothelial space in blood vessels to clotting factors in the blood, triggering the clotting cascade in the lumen of the blood vessel and forming an arterial blood clot called a thrombus, which occludes the artery in question. Thrombosis and embolism are the main differentiators in the causes of CHD. While a thrombus may remain stationary, it’s also prone to breaking off and travelling to other parts of the body through the bloodstream, thus becoming an embolus. This embolus can travel in the bloodstream until it occludes a blood vessel which is too small to let it through.

5 Ibid.
6 Ibid.
12 Ibid.
Coronary heart disease, alongside other cardiovascular diseases such as stroke and hypertension, are examples of non-communicable diseases. Non-communicable diseases, as opposed to communicable diseases, are not transmitted person to person\(^{13}\). Instead, they share several common risk factors “such as tobacco use, physical inactivity, alcohol consumption, and unhealthy diets”\(^{14}\). Risk factors which are specific to CHD also include high blood pressure, high cholesterol and family history. Given their non-transmittable nature, the increasing prevalence of non-communicable diseases can’t be controlled by treatments for the prevention of communicable diseases (e.g vaccines), thus making non-communicable diseases one of the most prominent causes of death globally as they are responsible for the deaths of 41 million people annually\(^{15}\). Instead, the primary treatments for CHD include lifestyle changes to mitigate the effects of its risk factors (e.g. diet changes, exercise and fitness), medication (e.g. to reduce blood pressure), and different procedures and surgeries to improve blood flow to the heart\(^{16}\) (e.g. coronary angioplasty, a procedure in which a stent is placed to open up occluded arteries\(^{17}\) ).

In this investigation, the Human Development Index is used as an indicator of socioeconomic development\(^{18}\). HDI data is composed of three indices to measure socioeconomic development: a health index (measured by life expectancy at birth), an education index (measured by mean and expected years of schooling), and an income index (measured by gross national income per capita)\(^{19}\). Performance in each index is expressed as a number from 0 to 1; the higher the number, the higher the level of socioeconomic status of a country. According to HDI data, countries can be classified as either developing or developed based on a cut-off HDI value: “countries with an HDI score higher than 0.788 are considered to be developed while countries with an HDI value lower than 0.788 are considered to be developing\(^{20}\).”

This investigation is particularly relevant to the contemporary healthcare industry. The gradual increase in mortalities caused by CHD over the last couple of years, as well as its prominence as the leading cause of death globally, have made the disease a major health threat, profoundly challenging public health sectors worldwide\(^{21}\). Non-communicable diseases such as CHD were once perceived as only being an issue for high-income countries. However, due to the rapidly changing lifestyles in low-middle income countries, such as societal change and urbanization\(^{22}\), non-communicable diseases have evolved to become a disproportionate burden in developing countries, where almost 80% of global deaths due to non-communicable diseases occur\(^{23}\). Ultimately, this investigation will potentially provide insight into how socioeconomic development may have consequences on CHD mortality, and what the implications hereof may be.

---

15 Ibid.
19 ABBB. 2013. Human Development Index and its components.
HYPOTHESIS

Hypothesis: there will be a negative correlation between HDI and the mortality rate due to coronary heart disease per 100,000 people in both developed and developing countries

A correlation can be predicted between HDI and the mortality rate due to CHD in both developed and developing countries. As outlined by Tandon et al. in their study, “an increased HDI of a country usually correlates with a more advanced healthcare system and greater access to healthcare.”24 Such a correlation would suggest that CHD patients would have a higher recovery rate in countries with higher HDIs given that the healthcare systems in these countries would be more equipped and able to treat CHD complications (e.g. by providing coronary angioplasties, appropriate medication)

APPROACH TO THE RESEARCH QUESTION

In order to answer the research question, I needed to develop a methodology which would provide me with accurate and precise results. Given the epidemiological nature of this investigation, I knew that I had to select two variables (one relating to HDI and the other relating to mortality rates due to CHD) and determine the correlation between them using data from different countries. One consideration I made when developing my methodology was that I needed to find a way to standardize the mortality rate data across the different countries I investigated, given that the number of people who die from CHD in a country is relative to that country’s total population. Given this, I decided that I would process my collected data in order to present the mortality rate due to CHD per 100,000 people, thus giving me a standardized way to compare different country’s data.

Originally, I had considered approaching the research question by plotting the mortality rate values against the HDI values for different countries in the same year. However, I realized that this approach may lead to erroneous results given that sampling errors and inaccuracies in data from that year in a particular country could not be taken into account. Consequently, I decided that I would take the mean mortality rate values and mean HDI values for different countries across several years in order to minimize such errors and inaccuracies in my investigation.

Lastly, I also considered ways to control confounding variables in my experiment. Confounding variables are external variables which are not the independent variable but may also have an effect on the dependent variable25. In this investigation, confounding variables are variables other than HDI that could influence the mortality rate due to CHD in particular countries. Given that confounding variables increase variance and introduce bias into an investigation, it is necessary to try and control them as much as possible. In order to do so I decided to develop an inclusion criteria in this investigation which would aim to exclude certain countries on the basis of specific attributes which countries must have to be chosen. Ultimately, this inclusion criteria will help enhance the accuracy of my results.

DATA SOURCES

The following databases were chosen to supply the data for this investigation on the basis of their credibility and reliability as data sources. The World Health Organization, United Nations Development Programme and the World Bank are all recognized as trustworthy, international institutions, and were thus chosen as suitable sources for the following data:

→ World Health Organization database on mortality due to CHD (ischaemic heart disease)26
→ United Nations Development Programme HDI data27
→ the World Bank population database per country28

VARIABLES

The independent variable in this investigation is HDI while the dependent variable is the mortality rate due to CHD per 100,000 people. It is also important to consider confounding variables in this investigation, such as “gender, age, dietary and lifestyle habits, smoking habits, diabetes and income inequality.”29 In order to control these confounding variables, an inclusion criteria was constructed.

INCLUSION CRITERIA

When constructing the inclusion criteria, it was necessary to take into account as many confounding variables in the investigation and control them in order to increase the reliability and accuracy of the results. One consideration to make is the location of the countries whose data is sampled. Given the many environmental risk factors of CHD, it was necessary to

24 Tandon, Ajay, Christopher Murray, Jeremy Lauer, and David Evans. 2014. MEASURING OVERALL HEALTH SYSTEM PERFORMANCE FOR 191 COUNTRIES.
filter the sample data such that the effects of confounding variables such as lifestyle and dietary habits are minimized. As such, I have chosen to limit the scope of my investigation to European countries, as defined by the World Health Organization30.

Another consideration which should be made in the inclusion criteria is the population of the countries whose data is sampled. Choosing a suitable population size is crucial for the investigation given that a sample size that is too small won’t provide a valuable insight into the relationship between HDI and mortality rates due to CHD. In the study by Zhu et al., it was suggested that “mortality rates from common conditions and diseases such as stroke and coronary heart diseases should be sampled from populations greater than 2 million”31. As such, a variable in the inclusion criteria for this investigation was chosen to be population, and chosen countries should have a population of at least 2 million people as of the year 2000. This effectively filtered out small European countries such as Iceland, Liechtenstein and Luxembourg.

Lastly, it was necessary to further group the eligible European countries in order to minimize the effect of confounding variables such as economic and cultural status which may impact the data collected in this investigation. The “Organisation for Economic Co-operation and Development” (OECD) was chosen to group the developed countries. The OECD is an economic organization of member states which share similar views on economic and political structure and also have similar socioeconomic statuses32. Ultimately, this should translate into similar lifestyle and dietary norms in these nations. The “Countries with Economies in Transition” (CEIT), otherwise known as the “Eastern European Group”, was used to group the developing countries. The CEIT is an official regional group in the United Nations of European countries with similar socioeconomic statuses. This regional group mainly consists of countries in the Balkans, Baltic and Caucasus, where the majority of developing countries in Europe are located. Taking the above-mentioned considerations into account, the final inclusion criteria was devised and is outlined in Table 1:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Inclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Developing countries</td>
</tr>
<tr>
<td>Location</td>
<td>Europe (as defined by the World Health Organization)</td>
</tr>
<tr>
<td>Population</td>
<td>&gt; 2 million since 2000 (as suggested by Zhu et al.)</td>
</tr>
<tr>
<td>HDI</td>
<td>&lt; 0.788 since 2000</td>
</tr>
<tr>
<td>Socioeconomic organization</td>
<td>Member of the CEIT</td>
</tr>
</tbody>
</table>

SAFETY, ENVIRONMENTAL AND ETHICAL CONSIDERATIONS
Given the nature of this investigation, there are no substantial safety or environmental considerations to be made. One minor thing to note is that all data should be used ethically and in accordance with the guidelines set by the database sources. These include, among other things, abiding to copyright procedures and the commercial use of the data provided.

METHODOLOGY AND TRIAL INVESTIGATION
Prior to conducting the final investigation, I conducted a trial investigation to determine whether a correlation could exist between HDI and mortality rates due to CHD. This trial investigation will also allow me to explain the methodology used to process the collected data. In order to ensure that the trial investigation was a fair representation of the entire sampling population, random sampling was used in order to randomly pick 5 developed and 5 developing countries. This was done using a random generator online33. The following countries were selected:

**Developing countries:** Albania, Azerbaijan, Croatia, Romania, The FYR of Macedonia

**Developed countries:** Czechia, Ireland, Italy, the Netherlands, the United Kingdom

To begin, data regarding the HDI, deaths due to CHD in thousands and total population in thousands was obtained for the selected countries and organized in Table 2 below. This data was collected across 4 separate years; 2000, 2010, 2015 and 2016. The particular years (2000, 2010, 2015 and 2016) were chosen because they were the only years when data regarding mortality rates due to CHD was available on the WHO database. Accordingly, data for the HDI of these countries and the total population was chosen for these same years as well. As is further shown in Table 2, the HDI values obtained from the United Nations Development Programme database and the values for deaths due to CHD in thousands were given to 3 significant figures. The data for the total population was given to zero decimal places.


### Table 2: raw data collection showing the HDI, deaths due to CHD in ‘000s and the total population in ‘000s of 10 randomly selected countries for the years 2000, 2010, 2015 and 2016

<table>
<thead>
<tr>
<th>Country</th>
<th>HDI</th>
<th>Deaths due to CHD ('000s)</th>
<th>Total population ('000s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>0.669</td>
<td>0.741</td>
<td>0.776</td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>0.640</td>
<td>0.740</td>
<td>0.758</td>
</tr>
<tr>
<td>Croatia</td>
<td>0.750</td>
<td>0.783</td>
<td>0.786</td>
</tr>
<tr>
<td>Czechia</td>
<td>0.796</td>
<td>0.862</td>
<td>0.882</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.857</td>
<td>0.909</td>
<td>0.929</td>
</tr>
<tr>
<td>Italy</td>
<td>0.830</td>
<td>0.870</td>
<td>0.876</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.876</td>
<td>0.910</td>
<td>0.926</td>
</tr>
<tr>
<td>Romania</td>
<td>0.695</td>
<td>0.715</td>
<td>0.767</td>
</tr>
<tr>
<td>The F.Y.R of Macedonia</td>
<td>0.669</td>
<td>0.735</td>
<td>0.754</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.867</td>
<td>0.905</td>
<td>0.918</td>
</tr>
</tbody>
</table>

In order to standardize the mortality data, given that the number of deaths due to CHD is relative to the population of a country, it was chosen to determine the mortality rate due to CHD per 100,000 people in each country. This mortality rate was calculated for each sampling year by first calculating the mortality rate due to CHD per person and then multiplying that value by 100,000 to represent 100,000 people. Sample calculations of how the mortality rate due to CHD per 100,000 people was calculated are shown below. It is important to note that the value for the mortality rate due to CHD per 100,000 people in all countries was correct to 3 significant figures, given that the value for deaths due to CHD was also to 3 significant figures.

**Sample calculation: mortality rate due to CHD per person in Albania in 2010**

\[
mortality \ rate \ per \ person = \frac{\text{deaths due to CHD in '000s in 2010}}{\text{total population in '000s in 2010}} = \frac{10.3}{2941} = 0.00350 \text{ deaths due to CHD per person (correct to 3 s.f.)}
\]

**Sample calculation: mortality rate due to CHD per 100,000 people in Albania in 2010**

\[
mortality \ rate \ per \ 100,000 \ people = mortality \ rate \ per \ person \times 100,000 = 0.00350 \times 100,000 = 350 \text{ (correct to 3 s.f.)}
\]

The mortality rates due to CHD per person and per 100,000 people for each sampling year are shown in Table 3:

### Table 3: processed data showing the mortality rate due to CHD per person and the mortality rate due to CHD per 100,000 people for 10 randomly selected countries for the years 2000, 2010, 2015 and 2016

<table>
<thead>
<tr>
<th>Country</th>
<th>Mortality rate due to CHD per person</th>
<th>Mortality rate due to CHD per 100,000 people</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>0.00279</td>
<td>0.00350</td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>0.00251</td>
<td>0.00233</td>
</tr>
<tr>
<td>Croatia</td>
<td>0.00569</td>
<td>0.00580</td>
</tr>
<tr>
<td>Czechia</td>
<td>0.00337</td>
<td>0.00324</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.00196</td>
<td>0.00115</td>
</tr>
<tr>
<td>Italy</td>
<td>0.00215</td>
<td>0.00178</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.00171</td>
<td>0.00117</td>
</tr>
<tr>
<td>Romania</td>
<td>0.00353</td>
<td>0.00404</td>
</tr>
<tr>
<td>The F.Y.R of Macedonia</td>
<td>0.00256</td>
<td>0.00256</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.00235</td>
<td>0.00147</td>
</tr>
</tbody>
</table>
Lastly, it was necessary to calculate both the mean HDI and the mean mortality rate due to CHD per 100,000 people by using the collected data over the four separate sampling years. Once again, the mean HDI values and the mean values for the mortality rate due to CHD per 100,000 people are both correct to 3 significant figures. Sample calculations for the two means are shown below.

**Sample calculation: mean HDI for Albania**

\[
mean \text{ HDI} = \frac{\text{sum of HDI values from different years}}{\text{number of different years}} = \frac{0.669 + 0.741 + 0.776 + 0.782}{4} = 0.742 \text{ (correct to 3 s.f.)}
\]

**Sample calculation: mean mortality rate due to CHD per 100,000 people in Albania in 2010**

\[
mean \text{ mortality rate due to CHD per 100,000 people} = \frac{\text{sum of mean mortality rates due to CHD per 100,000 people from different years}}{\text{number of different years}} = \frac{279 + 350 + 390 + 390}{4} = 352 \text{ (correct to 3 s.f.)}
\]

The mean HDI values and the mean values for the mortality rate due to CHD per 100,000 people for each country are shown in Table 4:

**Table 4:** processed data showing the mean HDI and mean mortality rates due to CHD per 100,000 people for 10 randomly selected countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Mean HDI</th>
<th>Mean mortality rate due to CHD per 100,000 people</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>0.742</td>
<td>352</td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>0.724</td>
<td>240</td>
</tr>
<tr>
<td>Croatia</td>
<td>0.777</td>
<td>572</td>
</tr>
<tr>
<td>Czechia</td>
<td>0.856</td>
<td>319</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.907</td>
<td>133</td>
</tr>
<tr>
<td>Italy</td>
<td>0.864</td>
<td>190</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.910</td>
<td>128</td>
</tr>
<tr>
<td>Romania</td>
<td>0.740</td>
<td>399</td>
</tr>
<tr>
<td>The F.Y.R of Macedonia</td>
<td>0.729</td>
<td>264</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.903</td>
<td>160</td>
</tr>
</tbody>
</table>

Ultimately, this data could be plotted on a scatter graph with the mean HDI values on the x-axis and the mean values for the mortality rate due to CHD per 100,000 people on the y-axis. Three different scatter graphs were created; one including all the countries and two other ones which were exclusively for the developing (Albania, Azerbaijan, Croatia, Czechia, the FYR of Macedonia) and developed countries (Ireland, Italy, Netherlands, Romania and the United Kingdom) identified in the trial investigation. Each graph includes a suitable trendline and a coefficient of determination (R²) value to evaluate the scatter of the data around the fitted trendline. The scatter graphs are shown in Graphs 1, 2 and 3 below:

---

**Graph 1:** scatter graph illustrating the relationship between HDI and mortality rate due to CHD per 100,000 people in all countries

**Graph 2:** scatter graph illustrating the relationship between HDI and mortality rate due to CHD per 100,000 people in developing countries

**Graph 3:** scatter graph illustrating the relationship between HDI and mortality rate due to CHD per 100,000 people in developed countries

The scatter plots in Graphs 1, 2 and 3 demonstrate correlations of varying strengths. As seen by the relatively high $R^2$ values (0.9566 and 0.7550), the positive correlation in the data from the developing countries (Graph 2) and the negative correlation in the data from the developed countries (Graph 3) seem to be valid. On the other hand, the correlation present in all countries (Graph 1) is not as strongly supported due to the relatively low $R^2$ value (0.3782). Despite the weak correlation shown in Graph 1, it was nonetheless decided to continue with the final investigation for all the countries that adhered to the inclusion criteria.

**INVESTIGATION AND RESULTS**

Using the same methodology as shown in the trial investigation, the data for all the countries which adhered to the inclusion criteria was collected and processed. No random sampling was required for the final investigation given that the inclusion criteria reduced the number of countries which were viable for this study to 31 countries. 19 of these countries were developed countries and the other 12 were developing countries. The raw data for the full investigation is available upon request while the processed data is shown in Tables 5 and 6 below:
Ultimately, this processed data was also plotted on scatter graphs. Similarly to the trial investigation, three separate scatter graphs were created: one including all the countries and two which were exclusively for the developing and developed countries. These scatter graphs are shown in Graphs 4, 5 and 6 below:

**Graph 4:** scatter graph illustrating the relationship between HDI and mortality rate due to CHD per 100,000 people in all countries

**Graph 5:** scatter graph illustrating the relationship between HDI and mortality rate due to CHD per 100,000 people in developing countries
STATISTICAL TESTING

When considering Graphs 4, 5 and 6, it is evident that a loose linear correlation exists between HDI and the mortality rate data. This linear correlation can be analyzed using a Pearson’s correlation in order to determine if the correlations observed are statistically significant. The Pearson’s correlation is a statistical test which measures the strength of a linear correlation between two continuous variables. The Pearson’s correlation coefficient (denoted by \( r \)) reveals both the strength (ranging from 0 to 1) and the direction (positive or negative value) of a linear correlation. Before conducting the Pearson’s correlation, it is necessary to consider the data requirements for the results of the test to be significant. Furthermore, the data used in a Pearson’s correlation must also be normally distributed given that the test is sensitive to outliers and skewness in the data. An effective way of analyzing the distribution of the data is by conducting a skewness analysis in which the skewness coefficients and their standard errors are determined for the two sets of data investigated. An example of a skewness analysis for all countries is shown in Table 7.

### Table 7: A skewness analysis of the HDI values and mortality rate data used in the investigation

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Skewness Coefficient</th>
<th>Standard error</th>
<th>Standard error × 2</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDI data</td>
<td>31</td>
<td>-0.288</td>
<td>0.498</td>
<td>0.996</td>
<td>0.288 &lt; 0.996</td>
</tr>
<tr>
<td>mortality rate data</td>
<td>31</td>
<td>0.837</td>
<td></td>
<td>0.996</td>
<td>0.837 &lt; 0.996</td>
</tr>
</tbody>
</table>

In order to determine whether the skewness coefficients aren’t too large to warrant concern of skewness, the absolute value of the skewness coefficients are compared to twice the value of their standard error. If the value for the skewness coefficient is less than twice its standard error, then there is no concern of skewness in the data and the Pearson correlation can be conducted. As is evident in Table 7, the absolute values of the skewness coefficients for both the HDI data and mortality rate data are smaller than twice their standard error, meaning there is no concern of skewness. Such a skewness analysis was also conducted for the data of the developing and developed countries independently (available upon request). In both cases there also wasn’t any concern of skewness. Given this, it was possible for a Pearson’s correlation to be conducted on all three datasets.

The Pearson’s correlation utilizes the following equation (equation 1) to determine the value of the correlation coefficient \( r \). In this equation, \( x \) can represent the HDI values, \( y \) can represent the values for mortality rate data due to CHD per 100,000 people and \( n \) represents the number of data pairs in the dataset (which in this case is 31).
**Table 7:** outline of the results of the Pearson’s correlation test for all countries, developing countries and developed countries

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Degrees of freedom</th>
<th>Level of confidence</th>
<th>Pearson’s r value</th>
<th>Critical value</th>
<th>Hypothesis test</th>
</tr>
</thead>
<tbody>
<tr>
<td>All countries</td>
<td>29</td>
<td>0.05</td>
<td>-0.658</td>
<td>0.349</td>
<td>H₀ rejected; a moderate correlation exists</td>
</tr>
<tr>
<td>Developing countries</td>
<td>17</td>
<td>0.05</td>
<td>0.695</td>
<td>0.456</td>
<td>H₀ rejected; a moderate correlation exists</td>
</tr>
<tr>
<td>Developed countries</td>
<td>10</td>
<td>0.05</td>
<td>-0.560</td>
<td>0.553</td>
<td>H₀ rejected; a moderate correlation exists</td>
</tr>
</tbody>
</table>

---


ANALYSIS AND CONCLUSION

Ultimately, this investigation has allowed me to answer the research question; “What is the correlation between HDI and mortality rate due to coronary heart disease (CHD) per 100,000 people and how does this correlation compare between developing (HDI > 0.788) and developed (HDI < 0.788) countries in Europe?” As is evident, the same correlations shown in the trial investigation (Graphs 1, 2 and 3) were also apparent in the final investigation (Graph 4, 5 and 6). However, the correlations in the trial and final investigation have different degrees of scattering around the fitted trendline, as is shown by the different R² values. While the correlation for all countries appears to be less scattered in the final investigation than in the trial investigation (0.4459 > 0.3782), the correlations for the developing and developed countries appear more scattered in the final investigation than in the trial investigation (0.5539 < 0.9956 for developing countries and 0.3142 < 0.7550 for developed countries). The final investigation suggested an overall negative correlation between HDI and mortality rate due to CHD per 100,000 people. In developed countries, this correlation was also negative. However, this appeared positive in developing countries. The statistical significance of these correlations is supported by the Pearson’s correlation test conducted. For all three instances, a moderate correlation was shown between HDI and mortality rates due to CHD per 100,000 people. A negative correlation was confirmed for the data in all countries (r = -0.658) and developed countries (r = -0.560) whereas a positive correlation was confirmed for the developing countries (r = 0.695).

Although the correlations found are supported by statistical testing, it is evident that the spread of data around the trendlines formed is not ideal. This is not only represented by the relatively low R² values in Graphs 4, 5 and 6, but also by the variation in the data collected. An example of this is seen in Table 6 in the variation of the values for mortality rate due to CHD per 100,000 people between Albania, Georgia, Romania and Turkey. While the mean HDI values of these 4 countries are relatively similar (ranging from 0.739 to 0.742), they have a large variation in the mean values for mortality rate due to CHD per 100,000 people (ranging from 216.9 to 311.4). Ultimately, this suggests that certain confounding variables have had an effect on the mortality rate due to CHD in particular countries, thus potentially skewing the results of the investigation. Moreover, the variation between these countries hints at methodological issues involved in the investigation, which could be improved to obtain more accurate results.

The results of the investigation don’t support my initial hypothesis that there will be a negative correlation between HDI and the mortality rate due to coronary heart disease in both developed and developing countries. However, this does not necessarily refute the reasoning behind the initial hypothesis made. From further research, the findings of this investigation seem to reflect the changing lifestyle habits, dietary habits and access to healthcare services in different stages of socioeconomic development. As was explained before, risk factors of CHD include high blood pressure and high cholesterol, both of which are caused by unhealthy lifestyle habits (e.g. dietary habits, physical inactivity, tobacco use and alcohol consumption). A study by Wu et al. suggests that unhealthy lifestyle patterns similar to this “are negatively correlated with socioeconomic development in middle- and high-income countries but positively associated with socioeconomic development in low-income countries”43. This is most probably a consequence of the “increasing consumer culture and rapid urbanization”44 experienced in developing countries. Critically, this study by Wu et al. could provide an explanation for why the correlation between HDI and mortality rates due to CHD is positive in developing countries, but negative in developed countries. Moreover, a study by Müller-Nordhorn et al. further suggests that “health literacy and accessibility is positively associated with socioeconomic development in developed countries”45. This increased health literacy and accessibility would allow more people in developed countries to obtain necessary treatment for their CHD complications (e.g. in the form of a coronary angioplasty or necessary medication) which would, in turn, further support the negative correlation between HDI and mortality rates due to CHD in developed countries.

In conclusion, the results of this investigation are supported by the idea that “an increased HDI of a country correlates with a more advanced healthcare system and greater access to healthcare”46. However, the investigation also suggests that factors such as the development of lifestyle and dietary habits alongside the levels of health literacy in a country play critical roles in the association between HDI and mortality rates due to CHD.

EVALUATION OF ERRORS AND IMPROVEMENTS

There are certainly some strengths in the methodology used in this experiment. This includes, for example, the trial investigation conducted to determine whether a correlation between the chosen variables could exist as well as the averaging of HDI and mortality rate values over four different years to obtain more precise data. Moreover, the reliable and credible data sources used (e.g. World Health Organization database) further enhanced the accuracy of the results obtained from this investigation. Lastly, the thorough statistical testing conducted, which included both a test for skewness as well as

46 Tandon, Ajay, Christopher Murray, Jeremy Lauer, and David Evans. 2014. MEASURING OVERALL HEALTH SYSTEM PERFORMANCE FOR 191 COUNTRIES.
the Pearson’s correlation test, allowed me to confirm the statistical significance of the correlations found in this investigation. However, given its epidemiological nature, the data correlations demonstrated in this investigation are inevitably susceptible to “confounding, reverse causality and ecological fallacy”47, which may have led to spurious correlations. This is due to the large array of confounding variables as part of the investigation which are difficult to control, such as diet and tobacco use, which could have impacted the results obtained.

Moreover, most of the major errors in the investigation stem from the inclusion criteria. In hindsight, developing more stringent inclusion criteria would help improve the accuracy and reliability of the investigation’s results. For instance, in order to improve the accuracy of the data collected, it might be helpful to include a criterion in the selection criteria regarding the quality of the data collected. Such a criterion would ensure that the data sources used have good coverage (e.g. being an accurate representation of the whole sampling population) and accuracy (no significant over- or under-estimation of mortality data), thus minimizing the risk of investigating a spurious correlation. An example of an index to measure data quality would be the one developed by Bhalla et al., which grades different data sources according to their “completeness, accuracy and timeliness”48.

Another flaw in the inclusion criteria was the lack of controlled variables used to select specific countries for the investigation. The countries in the investigation varied quite dramatically, from urban countries such as the United Kingdom to more sparsely populated countries like Sweden. This has led to some variance in the data (evident from the low R² values). As an improvement, it may be useful to develop criteria in the inclusion criteria which would account other factors, such as population density, climate and GDP per capita which would help further standardize the countries used in the investigation, despite the fact that this may limit the sample size.

The last flaw in the inclusion criteria is the very small minimum population size imposed to filter out countries (in the inclusion criteria, a country needed to have a population of at least 2 million people in order to be considered for the investigation). While this population size was adopted from an external study by Zhu et al., the low population sizes analyzed could affect the reliability of the data. As the population size decreases, the collected data becomes less of an accurate representation of the wider population assessed. As an improvement, a larger minimum population size could be imposed in the inclusion criteria, say 5 million people, in order to minimize variance and bias in the collected data.

Lastly, there are also some methodological errors which should be considered. One disadvantage of averaging the HDI and mortality rate data in my methodology was that fluctuations and anomalous data in particular years could not be differentiated within the data for a single country. This has implications on, for example, the Pearson’s correlation test conducted, which is very sensitive to outliers in data. As an improvement, it may be useful to plot all of the pairs of HDI and mortality rates from every year (2000, 2010, 2015, 2016) and label them in a way that illustrates which pairs are from the same country in different time periods. This wouldn’t only increase the sampling size of the investigation, but also allow the investigation to reveal how the association between HDI and mortality rates due to CHD has changed over time. Furthermore, it would make it easier to identify anomalous data for particular countries in particular years which could subsequently be removed from the collected data. This would ultimately ensure more accurate and reliable results from the investigation.

EXTENSIONS
As is evident from the evaluation above, several errors in this investigation stem from the existence of confounding variables which weren’t controlled through the inclusion criteria. While it seems difficult to completely isolate the mortality data collected from such confounding variables, it may be possible (as an extension to this investigation) to consider the effects of these confounding variables on mortality rates due to CHD and take these effects into account during the analysis of the data. For example, the correlation between mortality rates due to CHD with both age and gender (which are confounding variables in this investigation) could be considered. If it is seen that these variables are weakly correlated with mortality rates due to CHD, it could be concluded that they don’t have a significant impact on the correlation between mortality rates due to CHD and HDI. However, if either of these variables (e.g. age) demonstrated a significant correlation with mortality rates due to CHD this could be taken into account by considering the age demographic of the different developing and developed countries in the investigation and thus suggesting whether age could’ve affected any of the correlations found. Another possible extension to this investigation is investigating the social burden of CHD alongside its mortality rate. This can be done by using indicators such as the average hospitalization time for CHD cases or the costs of operations and procedures relating to CHD. Another way to investigate the social burden of CHD in different countries is by considering the correlation of HDI with DALYs (disability-adjusted life years) lost due to CHD. The DALY is an effective measure of disease burden as it can illustrate the “number of years lost to ill-health due to CHD”49 and how this differs in different countries with different HDI values. Ultimately, investigating the social burden of CHD as an extension will allow this investigation to better gauge the social impact and significance of CHD in different countries.

LITERATURE
Sources:

AABB. 2013. *Human Development Index and its components.*


https://doi.org/10.1136/heart.88.3.222.


https://courses.lumenlearning.com/boundless-statistics/chapter/hypothesis-testing-correlations/.


https://www.nhs.uk/conditions/coronary-heart-disease/treatment/.


https://www.psi.org/health-area/non-communicable-diseases/#about.


Tandon, Ajay, Christopher Murray, Jeremy Lauer, and David Evans. 2014. *MEASURING OVERALL HEALTH SYSTEM PERFORMANCE FOR 191 COUNTRIES*.


Images:

**Figure 1:** *Atherosclerosis - Biology - AS Level*. September 11, 2015. *Atherosclerosis - Biology - AS Level*. https://hubpages.com/education/Atherosclerosis-Biology.