Aalto University
School of Science

Master's Programme in Mathematics and Operations Research

On the effect of socioeconomic conditions on population health

Modeling disease burden in OECD countries

Eetu Reijonen

Master's Thesis
2026

Aalto University
School of Science

Author Eetu Reijonen

Title On the effect of socioeconomic conditions on population health — Modeling disease burden in OECD countries

Degree programme Mathematics and Operations Research

Major Systems and Operations Research

Supervisor Prof. Pauliina Ilmonen

Advisor Prof. Pauliina Ilmonen

Date 10 March 2026 **Number of pages** 66+1 **Language** English

Abstract

Large reductions in disease burden and increases in life expectancy have been achieved around the world in the last decades. Substantial health disparities exist between regions, however, and increasing healthcare costs put pressure on even the wealthiest countries. Thus, to be able to further improve population health, it is important to understand what drives the differences in disease burden between countries.

In this thesis, we examine how wealth, education, inequality, and healthcare resources affect population health. More specifically, we predict the disease burden of OECD countries, measured by disability-adjusted life years, from GDP per capita, Gini index, education index, and healthcare spending. Furthermore, exploratory data analysis is conducted on the five variables, Spearman's correlations are calculated, and heuristic clustering is performed with traditional and robustified principal component analysis (PCA). Multiple regression with variable transformations is used for modeling.

Analysis of the individual variables shows that there has been favorable development. Countries have become wealthier, more educated, and healthier, although economic inequality has largely remained the same. Regrettably, large differences between countries exist, most notably in health and wealth. PCA-guided heuristic clustering suggests forming three clusters of the OECD countries; European and Asian high-income countries differ from middle-income countries, and the United States is different from all others. Multiple regression shows a clear negative association between logarithmic healthcare spending and disease burden. Education and inequality do not have a large effect on disease burden, independent of healthcare spending.

The nonlinear association between healthcare spending and disease burden suggests that the higher spending a country has, the more expensive it is to further reduce disease burden. This is in agreement with the literature and implies that more years of healthy life could be saved with the same money in the least wealthy countries. Further research includes explaining the remaining variation in disease burden, including more countries in the analysis, and breaking down healthcare costs as well as disease burden by cause.

Keywords cost-effectiveness, disability-adjusted life year, epidemiology, healthcare, multiple regression, principal component analysis

Aalto University
School of Science

Tekijä Eetu Reijonen	
Työn nimi Sosioekonomisten olojen vaikutuksesta kansanterveyteen — Tautitaakan mallintaminen OECD-maissa	
Koulutusohjelma Matematiikka ja systeemitieteet	
Pääaine Systeemianalyysi	
Työn valvoja Prof. Pauliina Ilmonen	
Työn ohjaaja Prof. Pauliina Ilmonen	
Päivämäärä 10.3.2026	**Sivumäärä** 66+1 **Kieli** englanti

Tiivistelmä

Elinajanodote on noussut ja tautitaakka on laskenut viime vuosikymmenten aikana ympäri maailmaa. Alueiden välillä on kuitenkin edelleen merkittäviä terveyseroja, ja kasvavat terveydenhuollon kustannukset rasittavat rikkaimpiakin maita. Niinpä kansanterveyden parantamiseksi on tärkeä ymmärtää, mikä selittää tautitaakan erot maiden välillä.

Tässä diplomityössä tarkastellaan varallisuuden, koulutuksen, taloudellisen epätasa-arvon ja terveydenhuollon vaikutuksia kansanterveyteen. Ennustamme haittapainotetuilla elinvuosilla mitattua tautitaakkaa OECD-maissa käyttäen selittävinä muuttujina asukaskohtaista BKT:ta, Gini-indeksiä, koulutusindeksiä ja terveydenhuollon kustannuksia. Aluksi muuttujia tarkastellaan hyödyntäen eksploratiivista data-analyysia, korrelaatioanalyysia ja heuristista ryvästämistä, joka tehdään sekä perinteisen että robustifioidun pääkomponenttianalyysin avulla. Mallinnukseen käytetään monimuuttujaregressiota muuttujamuunnoksilla.

Tulokset osoittavat, että monilla osa-alueilla on tapahtunut myönteistä kehitystä. Maiden varallisuus, koulutustaso ja terveys on parantunut, vaikkakin taloudellinen epätasa-arvo on pysynyt pitkälti samana. Valitettavasti maiden välillä on edelleen suuria eroja varallisuudessa ja terveydessä. Pääkomponenttiohjattu heuristinen ryvästys tukee kolmen ryppään muodostamista; Euroopan ja Aasian korkeatuloiset maat eroavat keskituloisista maista, ja Yhdysvallat eroaa kaikista muista. Monimuuttujaregressio paljastaa selkeän negatiivisen yhteyden logaritmisesti muunnettujen terveydenhuoltokustannusten ja tautitaakan välillä. Koulutustasolla ja epätasa-arvolla ei ole suurta vaikutusta tautitaakkaan, kun erot kustannuksissa huomioidaan.

Epälineaarinen yhteys kustannusten ja tautitaakan välillä osoittaa, että mitä rikkaampi maa on kyseessä, sitä kalliimpaa tautitaakan laskeminen vaikuttaa olevan. Löydös vastaa muiden tutkimusten tuloksia ja tarkoittaa, että tietyllä rahamäärällä voitaisiin pelastaa eniten terveitä elinvuosia köyhimmissä maissa. Jatkotutkimuksena voitaisiin tarkastella jäljelle jäänyttä vaihtelua tautitaakassa, ottaa analyysiin mukaan enemmän maita, sekä tutkia yksityiskohtaisemmin terveydenhuollon kustannuksia sekä tautitaakan osatekijöitä.

Avainsanat kustannustehokkuus, haittapainotettu elinvuosi, epidemiologia, terveydenhuolto, monimuuttujaregressio, pääkomponenttianalyysi

Esipuhe

Haluan kiittää Professori Pauliina Ilmosta kannustavasta, avuliaasta, kehittävästä ja sopivan vaativasta ohjauksesta, sekä mielenkiintoisesta aihe-ehdotuksesta. Minulla oli kunnia olla yksi hänen viimeisistä valvottavistaan (ainakin vähään aikaan) ennen siirtymistä uusiin tehtäviin. Toivotan hänelle onnea ja intoa uudessa roolissaan perustieteiden korkeakoulun dekaanina kesäkuusta alkaen.

Espoo, 13.2.2026

Eetu Reijonen

Contents

Abbreviations

AI	Artificial Intelligence
BMI	Body Mass Index
CE	Cost-Effectiveness
CMNN	Communicable, Maternal, Neonatal, and Nutritional (diseases)
CV	Coefficient of Variation
DALY	Disability-Adjusted Life Year
GBD	Global Burden of Disease
GDP	Gross Domestic Product
GNI	Gross National Income
EI	Education Index
HIV	Human Immunodeficiency Virus
ICD	International Classification for Disease
IHME	Institute for Health Metrics and Evaluation
IMF	International Monetary Fund
MCD	Minimum Covariance Determinant
NCD	Noncommunicable Disease
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PPP	Purchasing Power Parity
RCT	Randomized Controlled Trial
SF	Healthcare Spending as Fraction of GDP
SP	Healthcare Spending (in USD)
UN	United Nations
USD	United States Dollar
WHO	World Health Organization
YLD	Years Lived with Disability
YLL	Years of Life Lost

1 Introduction

Health is paramount to living a fulfilling life. Vast leaps in reducing disease and averting early death have been made in the last centuries. As a consequence of the advances in epidemiology, the health status of populations is better known than ever, which enables identifying and confronting the most serious health issues of our time [1]. Throughout the last few decades, global health has furthermore improved significantly, but large differences remain between regions of the world [2]. In addition to improving the standard of living, health interventions including general healthcare, vaccination programs, and prevention campaigns can be effectively used to reduce disease burden. However, aging populations, costs of advanced treatments, and the rise of noncommunicable diseases put a high burden on healthcare systems — while many governments are trying to decrease spending. Therefore, it is crucial to understand how changes in healthcare spending might affect population health, and to consider which other factors have an effect on disease burden. This can help people make more informed choices, and we can stay on the path of reducing death and suffering due to disease, for everyone around the world.

Goals and contributions This thesis examines disease burden and healthcare from a wide perspective. Our analysis seeks to uncover how much of the variation in disease burden between OECD countries is explained by wealth, education level, economic inequality, and healthcare spending. The primary goal is to see how health spending affects health outcomes between countries and whether there are other socioeconomic factors which explain the remaining differences. In addition to studies on the cost-effectiveness of specific health interventions (e.g. [3–11]), analysis similar to ours has been conducted where correlations between healthcare spending and disease burden are examined [12]. Our study adds education level and economic inequality into the analysis to see how they affect disease burden in addition to healthcare spending and GDP, and how the variables predict one another. The results can guide public policy to achieve healthier populations and more cost-effective healthcare spending.

Scope, constraints, methods Our analysis is limited to the OECD countries to have reliable and complete data. Several factors including economical, social, cultural and even geographical ones affect the disease burden in a country. We have chosen GDP per capita, education index, Gini index, and healthcare spending as the explanatory variables because they offer a broad summary of the living conditions of countries. GDP is a rough indicator for the standard of living and Gini summarizes economic inequality. To measure disease burden, we use the disability-adjusted life year -metric developed in the Global Burden of Disease (GBD) study (see Section 2.3.6). Univariate statistics, correlation analysis, and multivariate clustering are used for exploratory analysis and multiple regression is employed in studying the relationship between the explanatory variables and disease burden. Our results cannot establish causality but can be used as a starting point for generating hypotheses or for further study on the specific factors affecting disease burden.

Structure of the thesis The structure of the thesis is as follows: Section 2 gives the reader the necessary background to understand the work. It introduces the OECD countries, socioeconomic indicators, and topics in epidemiology. Section 3 discusses the current state of global health, surveys health interventions and their cost-effectiveness, and shows similar work to ours. The data sources and the data analysis methods are presented in Section 4. The results of the study are shown in Section 5. Section 6 discusses the implications of the results as well as the limitations of our analysis. Section 7 concludes the thesis. At the end, Appendix A describes the use of artificial intelligence (AI) tools.

2 Background

This section introduces the necessary concepts and topics that are required to understand the work. First, the OECD countries are presented, and the economic and social indicators that are used as explanatory variables in our analysis are introduced. Then the basics of epidemiology are discussed, including the measurement of disease burden which is central in our analysis. Finally, the Global Burden of Disease study is introduced.

2.1 Organisation for Economic Co-operation and Development

OECD is an organization of countries which aims to improve living standards through public policy

According to its official website, the Organisation for Economic Co-operation and Development (OECD) "is an international organisation that works to build better policies for better lives." The OECD was founded in 1961 as a continuation to the Organisation for European Economic Co-operation which administered North American reconstruction aid after World War II. Initially, 20 member countries formed the organization, but it currently consists of 38 countries with many more applying for membership. The OECD develops policy recommendations and standards, and conducts studies such as the PISA (education study). The 2025 budget of the OECD was 361.1 million EUR. [13]

The OECD countries include mostly high-income, democratic countries in Europe and North America as well as some countries from Asia, South America, and Oceania. The countries are regarded as developed, and they typically have very high Human Development Index (HDI). As of 2025, the member countries are Australia, Austria, Belgium, Canada, Chile, Colombia, Costa Rica, Czechia, Denmark, Estonia, Germany, Finland, France, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Slovenia, Slovakia, South Korea, Spain, Sweden, Switzerland, Turkey, the UK and the US [13].

The countries are shown in Figure 1. To increase the visibility of small countries, for the rest of the thesis, the countries will be shown in three separate maps: Europe, the Americas, and East Asia and Pacific. We have chosen to limit our analysis to the OECD countries to maximize data availability and reliability, and to have economically and socially similar countries to compare the effects of healthcare spending.

2.2 Socioeconomic indicators

2.2.1 Gini coefficient

Gini measures economic income inequality

Gini coefficient (or Gini index) is a summary of the income distribution in a country. It varies from 0 which represents perfect equality (everyone has the same income) to 1

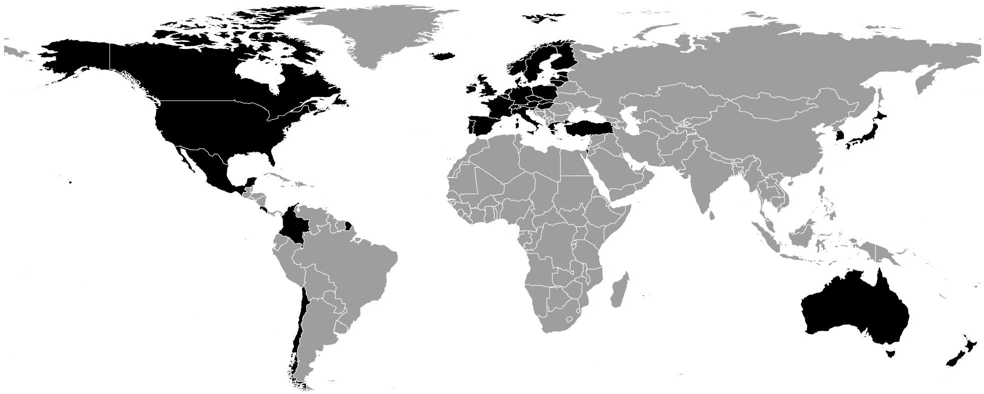

Figure 1: OECD countries as of December 2025, shown in black.

which represents perfect inequality (one person has 100% of the income). Alternatively, sometimes it is reported on a scale from 0 to 100. The income levels are estimated from nationally representative household surveys. Countries with Gini over 0.4 are classified as having high inequality. [14] As a reference, values of the Gini coefficient range typically between 0.2 and 0.4 in European countries.

Formally, the Gini coefficient is defined as the relative difference in the areas under the Lorenz curves of a perfectly equal income distribution and the actual income distribution of a country. A typical Lorenz curve is shown in Figure 2. Points on the curve correspond to the share of national income obtained by the lowest x% of people. For example, the poorest 20% might earn 10% of the national income. In a perfectly equal distribution, the bottom x% would earn x% of the income. A curve representing perfect income equality is thus a straight line (outlining area **A** in Figure 2). The actual income distribution always differs from perfect equality and is thus under the straight line (outlining area **B** in Figure 2). For a country with income distribution represented by Figure 2, the Gini coefficient would be calculated from the ratio of the areas, i.e., Gini $= A/(A + B)$.

Gini coefficient is not unique in the sense that different income distributions can produce the same value for the coefficient. It is also important to note that Gini can increase even when wealth increases for every person, because it is a measure of relative, not absolute income differences. Finally, the household surveys can differ in their methods (e.g. whether income or household consumption is measured) and the results between countries or even between years are not fully comparable. Despite the limitations, Gini is the most widely used metric for income inequality and allows for easy comparison between countries in different years. For more information about the Gini coefficient, its data sources, methodologies, and limitations, the reader is directed to the World Bank data portal [14] and the Poverty and Inequality Platform Methodology Handbook [16].

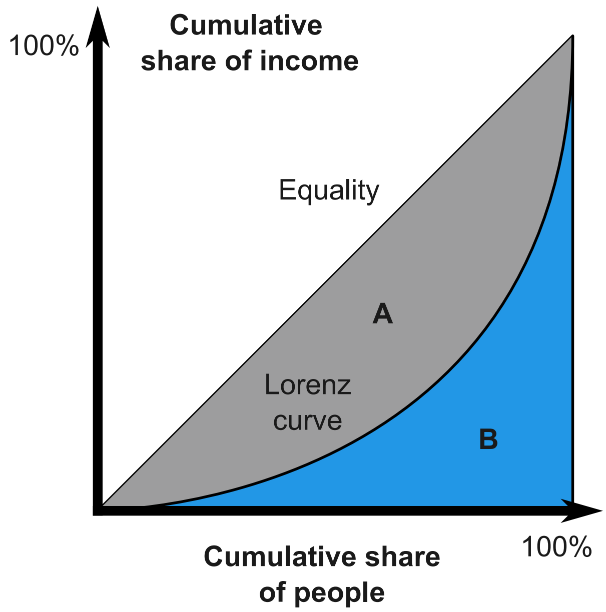

Figure 2: Lorenz curve represents the income distribution in a country. The visualization was introduced in [15].

2.2.2 Gross Domestic Product

GDP is the total amount of economic value created in a country within a time period

Gross domestic product (GDP) measures the economic value of the total amount of goods and services produced within a given area and time period. Most often it is calculated for a country, quarterly and annually. Everything that is produced inside a country is counted towards GDP regardless of ownership. It is the most commonly used measurement of economic output around the world. [17] The modern form of GDP was developed by Simon Kuznets for a United States government report in 1934 [18]. To assess the average level of income in a country, GDP is divided by the population size, which yields GDP per capita. Despite its limitations, GDP per capita can be seen as a rough indicator for the standard of living. According to the United Nations, in 2023, the highest GDP per capita was in Monaco at 256 581 USD while the lowest was in Yemen at 222 USD. Four countries had GDP per capita greater than 100 000 USD while over 20 countries had GDP per capita less than 1 000 USD, which shows that there are massive differences between nations in wealth. [19]

GDP can be calculated via different approaches. Here we introduce the expenditure approach, where the expenditures made by households, businesses and the government are considered. Usually, expenditure statistics are reliable and thus this method yields an accurate estimate of GDP. Using the expenditure approach, GDP can be calculated as:

$$\mathrm{GDP} = \mathrm{C} + \mathrm{I} + \mathrm{G} + \mathrm{NX},$$

where C is the consumption by households and non-profit organizations, I is business expenditures (investments) and new houses bought by households, G is the government spending, and NX is the net exports (exports minus imports). [17]

Important limitations to GDP are that it does not include unpaid household work or damage to the environment. Activities that hinder the ability to produce goods in the future, such as overfishing, can increase GDP. Likewise, activities that lower the quality of life, such as traffic jams in the form of wasted time and air pollution, can increase GDP, in this case in the form of increased fuel usage. Unpaid work, such as house cleaning, cooking, and caring for children, the sick, and the elderly, could be as high as 50% of a country's GDP [20]. Furthermore, part of the economic value produced in a country might be spent elsewhere. This is the case when many foreign companies operate within a country and pay their profits to foreign owners. Gross national income (GNI) is metric which takes this into account and thus might be a better measure of economic well-being [21]. Still, we chose to examine GDP due to its better familiarity to most readers.

Due to inflation, raw GDP values must be adjusted to reflect true economic growth. For example, if a country's GDP increases from 10 billion USD to 30 billion USD within a 20-year period but the value of money decreases by 50% (corresponding to an annual inflation rate of 3.5%), the real GDP increase is 50%, not 200%. The value before the adjustment is referred to as *nominal GDP* while *real GDP* is the value after the adjustment [22]. In addition, the purchasing power of money varies between countries due to differences in average wages, among other reasons. This is accounted for by purchasing power parity (PPP) adjustment. [23] For example, the cost of an equivalent meal or equivalent housing differs between countries. In our work, we use the real GDP values per capita, adjusted for purchasing power parity.

2.2.3 Education Index

Education index combines levels of mean schooling and expected schooling

The Human Development Index (HDI), developed by the United Nations (UN), is a combined metric for education, life expectancy, and national income. It is widely used in comparing the living conditions between countries and offers a more comprehensive alternative to GDP per capita. It is calculated as the geometric mean of three indices: index of logarithmic gross national income, education index (EI), and life expectancy index. [24]

Initially, we considered including HDI as one of the explanatory variables in our analysis but changed it to education index (EI). This is because GNI is very similar to GDP and life expectancy is tied to the calculation of disability-adjusted life years, the response variable in our analysis. Thus, the inclusion of HDI would create dependencies in the model which would hinder interpretability.

The education index (EI) varies between 0 and 1 and is calculated using the mean years of schooling (MS) for adults over 25 and the expected years of schooling (ES) for a child entering the educational system:

$$\mathrm{EI} = \frac{1}{2}\left(\frac{\min(15, \mathrm{MS})}{15} + \frac{\min(18, \mathrm{ES})}{18}\right).$$

The education index is maximized if MS and ES surpass 15 and 18 years, respectively. 18 years of education corresponds to achieving a master's degree — 10 years of primary education, 3 years of secondary education, and 5 years of higher education (university or equivalent).

2.3 Epidemiology

Epidemiology studies death and disease

Epidemiology is the fundamental science of public health. It studies the causes, distribution, and prevention of diseases — how population health can be improved. One of the earliest epidemiological studies was conducted by John Snow in London around 1850. He found an association between cholera outbreaks and certain public wells. The findings had an impact on public policy, water sources were improved, and consequently, the spread of cholera was limited before its cause (a bacterium) was identified. [1]

In epidemiology, the three most important areas of study are detecting the causal factors for diseases, describing population health, and evaluating the effectiveness of interventions. Many diseases are the result of multiple causes, including genetic, behavioral, and environmental factors. The latter are especially important in epidemiology as they can easily go unnoticed in traditional healthcare contexts. Describing population health includes the quantification of death, diseases, and injuries, and how they are distributed in a population. Understanding their changes is crucial for evaluating interventions (e.g. vaccines or reductions in pollution) and for finding the most effective ways to improve population health. [1]

Some major achievements of epidemiology include the elimination of smallpox through providing information about the spread of the disease and evaluating control measures, as well as showing the dangers of smoking and the subsequent reductions in tobacco consumption. The following Sections 2.3.1–2.3.4 give an introduction to epidemiology. They summarize the World Health Organization (WHO) book by Bonita et al. [1] which is an excellent resource for learning more about the topic.

2.3.1 Measuring disease

Defining what health is can be difficult. The definition proposed by the WHO in 1948 involves not merely the absence of disease but a state of well-being. For the purposes of epidemiology, however, definitions have to be chosen to quantify health and disease. Sometimes arbitrary cut-offs are chosen. For example, cardiovascular disease risk increases continuously with blood pressure, but there is a certain limit for "high blood pressure". For simplicity, in this thesis we talk about health and the absence of disease interchangeably.

Mortality Measuring death is often simpler than measuring disease. Many countries keep detailed statistics on the deaths and births in their population. The causes of death are often reported as well. Standards, such as the International Classification for Disease (ICD), exist to ensure that the assigned causes of death are commonly recognized. Data on deaths and their causes, however, is usually only available from high-income and middle-income countries. Sometimes verbal autopsy — an interview of the deceased person's family — is used for hypothesizing the cause of death. Thus, only estimates are often available from developing countries.

There are multiple ways to quantify mortality. The frequency of death is usually expressed in death rates (as the number of deaths per 1000, for example). The rates can be calculated by cause, age group, and gender. This allows for comparisons between the rate of fatal road accidents in 20-year-old men and women, for example. Another very common metric is life expectancy which is the expected number of years until death from birth, calculated using the death rates of each age group.

Morbidity Diagnostic criteria are used to determine the presence of an illness. They include symptoms, test results, and history (description of what has happened before the disease developed and how the symptoms have progressed).

Frequency of disease is quantified by incidence and prevalence. They are calculated for the population at risk which can include everyone or a subgroup of the population (e.g. only men are at risk of developing testicular cancer). Incidence is the number of new cases in a time period, usually a year, and prevalence is the number of existing cases at a given time. Thus, incidence can be interpreted as the probability of an individual becoming ill within a time period and prevalence as the probability of an individual being ill. As an example, the common cold has high incidence but low prevalence, while diabetes has low incidence but high prevalence.

Some diseases, such as most mental illnesses, have low mortality but still cause high disease burden. Thus, a measurement of morbidity is needed. The most commonly used metric is the disability-adjusted life year (DALY) which combines years of life lost due to early death and healthy years lost due to diminished quality of life. DALYs are introduced more thoroughly in Section 2.3.5.

2.3.2 Types of studies and inferring causality

Epidemiological studies are classified into two categories: observational and experimental. The former involve only describing and measuring phenomena or quantifying relationships between health outcomes and different factors, e.g. pollution or lifestyle. In experimental studies, health interventions are compared by assigning them to a group of subjects and comparing the outcomes to a group without the interventions. In many study designs, it is hard to prove causality, and confounding can lead to erroneous conclusions; carrying matches or a lighter is associated with lung cancer but does not cause it.

Observational studies Observational studies can be further divided into descriptive and analytical studies. As their name suggests, descriptive studies report the health status of a group, but they make no attempt in analyzing cause and effect. They are, however, useful in identifying health trends and prompting further research. In addition, they can be the first step in identifying a new disease. For example, the discovery of the human immunodeficiency virus (HIV) started from the case reports of a rare form of pneumonia.

Analytical observational studies include correlational, cross-sectional, case-control, and cohort studies. Correlational studies are the easiest to conduct and offer the broadest view on health phenomena. The data often already exists and does not need to be collected. Groups of people are compared statistically and associations are sought between a factor and an outcome. Correlational studies are useful for generating hypothesis but provide no proof of causation nor an explanation for the findings. This thesis is a correlational study: it searches for an association between healthcare spending, education, inequality, wealth, and disease burden. Similarly, the Global Burden of Disease study (see Section 2.3.6) — the largest global study on health — is both a descriptive and a correlational study as it both reports the death and disease statistics around the world, and investigates the associations between risk factors and diseases. The other study types are introduced in the WHO book [1].

Experimental studies Experimental studies provide the strongest evidence for causality, but they are the most difficult and expensive to conduct. Furthermore, there are often ethical challenges with them; assigning people randomly to smoking and non-smoking groups to study the risk of lung cancer is not acceptable. However, this study design, called a randomized controlled trial (RCT), is used in determining the effectiveness of all vaccines and medications. The health outcome of a group treated with the medication is compared with a group that has not been treated (and a group that is treated with a placebo, a harmless fake medication). Field trials and community trials are similar but in contrast to RCTs, these are not conducted in a controlled laboratory or hospital setting but "in the field".

Confounding Confounding refers to the effect of a hidden variable on an outcome. This can lead to inferring spurious relationships between variables. A well-known example is caffeine and cigarettes. Consumption of coffee seems to be associated with heart disease. However, people who consume more coffee also tend to consume more cigarettes (smoking causes heart disease). When comparing heart disease risk between non-smoking coffee drinkers and non-drinkers, there is no difference. Thus, cigarette smoking is a confounder. Ideally, when the effect of some factor on health outcomes between groups is measured, everything except that factor should be same across the groups. In addition to the study design, confounding can be controlled for by statistical methods, e.g., stratification where health outcomes are compared by age group.

2.3.3 Communicable and noncommunicable diseases

In epidemiology, an important distinction is made between communicable diseases and noncommunicable diseases (NCDs). Patterns of disease, and consequently, the necessary interventions are fundamentally different for these two categories. In the Global Burden of Disease study, causes of death and ill health are classified into injuries, noncommunicable diseases, and communicable, maternal, neonatal and nutritional (CMNN) diseases. Diseases in the last category are typically associated with poverty and inadequate living conditions.

Noncommunicable diseases As their name implies, noncommunicable diseases are diseases that cannot be transmitted from one person to another (or between humans and animals). The most common noncommunicable diseases include cardiovascular diseases, cancer, chronic respiratory diseases, diabetes, and mental health disorders. They account for the majority of disease burden in the world, especially in high-income countries. In fact, most of the progress in reducing disease burden globally has come from mitigating communicable diseases while burden from noncommunicable causes has decreased much less. The treatment of noncommunicable diseases is often difficult and very expensive, such as is the case with cancers. Encouragingly, however, much of communicable disease burden can be prevented via interventions targeting people who are not yet sick.

There are four levels of prevention. Primordial prevention aims to improve the social, economic, and environmental conditions which contribute to disease risk. Examples include smoking, physical activity, healthy diet, and pollution. Primary prevention controls specific risk factors, e.g. high blood cholesterol. In secondary prevention, consequences of diseases are limited with early detection and treatment. Screening is used to identify individuals or groups at risk. For example, cervical cancer screenings and testing sight in children are effective ways to start treatment early. Tertiary prevention includes most traditional medical care. Its goal is to limit the progression of or the complications from an existing disease. Among the four levels, most can be accomplished in the primordial phase with the least effort.

Communicable diseases Communicable diseases are caused by an infection from a pathogenic agent. These include, bacteria, viruses, fungi, and protozoa. Some communicable diseases are contagious, meaning that they spread between humans with via direct contact. Others are transmitted indirectly through vehicles (e.g. contaminated water) or vectors (e.g. insects or animals). The communicable diseases that cause the most deaths around the world are acute respiratory diseases (e.g. pneumonia), HIV, diarrheal diseases, tuberculosis, malaria, and measles. Communicable diseases cause high disease burden in low-income countries, but respiratory diseases, especially, also cause many deaths in high-income countries, particularly among the young and the elderly. The COVID-19 pandemic has shown that infectious diseases can still be a large threat to collective health, even in high-income countries. The most important ways of preventing communicable diseases include immunization (vaccines) and improving living conditions.

2.3.4 Health policy

Health policy is the set of decisions that enable the achievement of certain health goals in society. Population health is not only influenced by policies related to the healthcare sector; everything from environmental regulations to food taxation can have a large impact (often times even larger). Epidemiology can and does influence health policy. Frequently, its effects are mediated via public opinion. Frustratingly, decision-makers often ignore the evidence provided by epidemiology, apply it too late, or even act contrary to it. The overall political climate can have large repercussions. The current health administration of the United States, for instance, promotes vaccine skepticism and the consumption of saturated fat — two positions which could not be in more disagreement with scientific evidence. A major way to shift public opinion and inform policymakers is through the release and dissemination of information. For example, the *Disease Control Priorities* -books released by the World Bank Group detail what should be done to most effectively reduce global disease burden [25].

The health planning process is a systematic approach for intervening in a specific disease. It involves assessing the burden, understanding the causes, evaluating the effectiveness of interventions, and finally implementing them. The previous sections have already discussed the first two phases. The effectiveness of interventions is usually evaluated with cost-effectiveness (CE) analysis or cost-benefit analysis. CE analysis compares costs with health outcomes: how many dollars does it take to save one year of healthy life, for instance. Cost-benefit analysis also assigns a monetary value to diseases (healthcare costs, lost productivity, and compensation). Section 3 reviews the literature analyzing the cost-effectiveness of interventions.

2.3.5 Disability-Adjusted Life Year

Disease burden is measured by DALYs

The disability-adjusted life year (DALY) is a combined metric for disease burden, injuries, and death within a population. In simple terms, it measures the number of healthy years of life lost. DALY is defined as the sum of years of life lost (YLL) and years lived with disability (YLD), i.e.

$$\text{DALY} = \text{YLL} + \text{YLD}.$$

Figure 3 shows a visual explanation of the DALY metric. The metric was introduced in the first Global Burden of Disease Study in 1990 [26]. It was developed to provide a more comprehensive view of population health compared to death statistics, and to enable comparisons between different populations. The following is an intuitive introduction to DALYs. A lengthier introduction can be found in [27] and the mathematical methods for the most recent GBD 2023 study in [28, Supplementary Appendix 1].

YLLs For an individual, years of life lost is the remaining life expectancy at death. For example, someone who dies at twenty years old loses about 66 years of life while

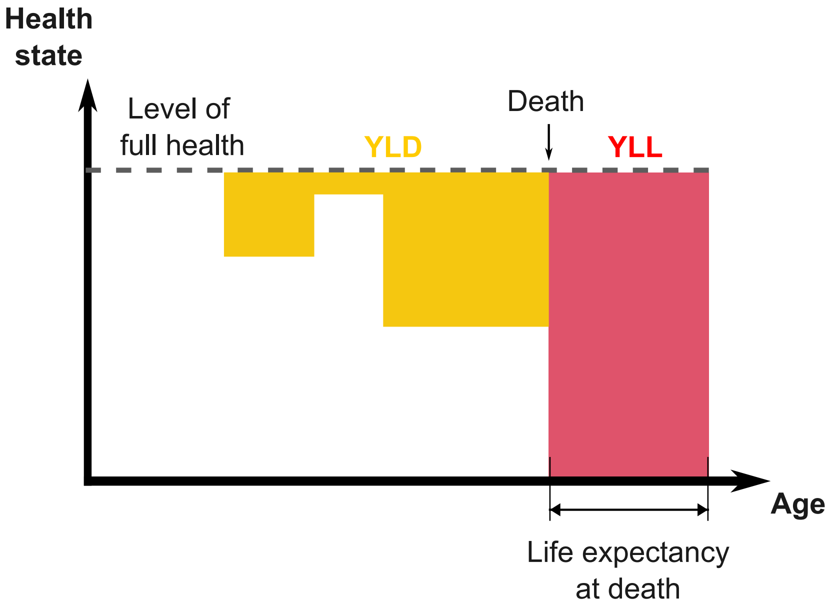

Figure 3: A visual explanation of the DALY metric. The total area under the dashed line represents a person's expected life span in full health. YLLs and YLDs take away from that ideal life. Conceptually inspired by [29].

a 70-year-old loses 17 years. The method for calculating the remaining life expectancy has varied between studies and institutions but the most recent GBD studies have "used the lowest observed age-specific mortality rates by location and sex across all estimation years from locations with total populations greater than 5 million in the most recent year of estimation to establish a theoretical minimum risk reference life table" [28, Supplementary Appendix 1, p. 46]. Thus, the remaining life expectancy is the best global empirical life expectancy for someone at the given age. For a population, the YLLs are calculated by multiplying the number of deaths at each age with the remaining life expectancy at that age. Thus, YLLs are straightforward to calculate and rely on relatively reliable statistics as only the years of birth and death have to be recorded. As an example, Table 1 presents the remaining life expectancies used in GBD 2010.

YLDs Compared to YLLs, the calculation of years lived with disability is more complicated. YLDs represent the amount of time a person lives in less-than-perfect health, and the discomfort associated with it. On a population level, YLDs are calculated using the prevalence of diseases, but the metric is easier to understand on an individual level. For an individual, YLDs are calculated by multiplying the time of having a disease or a disability by its severity which varies between 0 (perfect health) and 1 (state equivalent to death). For instance, someone suffering from moderate hearing loss (0.027 disability weight) for 20 years will accumulate $0.027 \cdot 20 = 0.54$ YLDs from this ailment. A more serious condition such as severe dementia (disability

Table 1: The remaining life expectancy table used in GDB 2010. Life expectancies for ages 85+ are not shown separately in [27] although the age group 85+ is further broken down in GBD 2010. Adapted from [27].

Age range	Life expectancy	Age range	Life expectancy
Neonatal	86.01	40–44	44.23
Postneonatal	85.68	45–49	39.43
1–4	83.63	50–54	34.72
5–9	78.76	55–59	30.10
10–14	73.79	60–64	25.55
15–19	68.83	65–69	21.12
20–24	63.88	70–74	16.78
25–29	58.94	75–79	12.85
30–34	54.00	80–84	9.34
35–39	49.09	85+	5.05

weight 0.449), will cause a person to accrue the same YLDs in less than two years living with the disease.

Table 2 shows the disability weights for some diseases that were used in GBD 2019. The disability weights are obtained from global surveys of tens of thousands of people from different cultures where laypersons (normal people, not medical professionals) are asked to choose the worse option from two health conditions based on short descriptions. The weights are obtained from the rankings with probit regression analysis. [30] Example lay descriptions are shown in Table 3.

Table 2: Disability weights for some diseases in the GBD 2019 study. Higher values indicate that the disease has been ranked as more severe by the surveyed.

Disease	Disability weight
Asthma: uncontrolled	0.133
Infectious disease: acute episode, mild	0.006
Cancer: metastatic	0.451
Alcohol use disorder: moderate	0.373
Low back pain, moderate	0.054
Concussion	0.214

Age-adjustment Finally, to enable comparisons, total DALYs are population-corrected and age-adjusted. The corrections are important because without them, the raw DALY numbers do not allow for comparisons between countries or even between years within a country with a changing age distribution or population size. The resulting metric is called the age-adjusted DALY rate, and it is presented as DALYs per 100 000. The values ranged from 17 541 to 30 773 in the OECD countries in 2023. Population correction and age-adjustment are accomplished by calculating the DALY

Table 3: Disease descriptions used in determining the disability weights for the GBD 2015 study. From [27].

Disease	Description
Asthma: uncontrolled	has wheezing, cough and shortness of breath more than twice a week, which causes difficulty with daily activities and sometimes wakes the person at night.
Infectious disease: acute episode, mild	has a low fever and mild discomfort, but no difficulty with daily activities.
Cancer: metastatic	has severe pain, extreme fatigue, weight loss and high anxiety.
Alcohol use disorder: moderate	drinks a lot, gets drunk almost every week and has great difficulty controlling the urge to drink. Drinking and recovering cause great difficulty in daily activities, sleep loss, and fatigue.
Low back pain, moderate	has moderate back pain, which causes difficulty dressing, sitting, standing, walking, and lifting things.
Concussion	has headaches, dizziness, nausea and difficulty concentrating.

rates for each age group and taking the weighted average using a standard population [31, Supplementary Appendix 1].

2.3.6 Global Burden of Disease

The GBD study is a biennial effort to quantify world health

The Global Burden of Disease Study is the largest and most comprehensive research project which attempts to investigate and compare the prevalence and the causes of death and disease around the world. The results are published biennially. In addition to the estimates for the prevalence of diseases and death, GBD creates projections for future disease burden [32] and provides its data to the public and to other researchers, sparking a vast number of more specific analyses related to certain causes or geographical areas, e.g. [33, 34].

Started in 1990 and commissioned by the World Bank, the first Global Burden of Disease study reported death and disease in 8 regions for 106 different causes and had just two authors [26]. Since then, the research has greatly expanded, reporting the disease burden for 204 countries and 371 different causes. Now each article has an average of 242 authors and there are thousands of collaborators around the world. [35] Since 2010, GBD has been coordinated by the Institute for Health Metrics and Evaluation (IHME) at the University of Washington. The funding is provided by the Gates Foundation. A summary of the history of GBD can be found in [27].

Throughout its history, GBD has facilitated the development of many statistical methods and methodological improvements in descriptive epidemiology to deal with missing or conflicting data, differing case definitions, and biased reporting. The results of GBD are used for guiding national policymakers in many countries, and the development of the DALY metric has enabled the cost-effectiveness analysis of health interventions. According to Murray, the initiator of GBD, the future of the project entails making the results more accessible, transparent, and relevant to decision-makers and the public. To facilitate the utilization of the results in policy formation, more local collaboration is sought. Lastly, GBD attempts to answer the demand for more forecasts as well as providing real-time data and analysis. [35]

3 Literature review

In this section, the current state of global health and recent developments are summarized from the latest 2023 Global Burden of Disease study. In addition, literature linking the effects of healthcare spending and some health interventions to health outcomes is presented along with some cost-effectiveness studies. Finally, similar work to ours is briefly discussed.

3.1 State of global health

Health has improved around the world, but there are vast differences between regions

Since 1990, disease burden has decreased and life expectancy has increased almost everywhere. The greatest health improvements have been made in the poorest countries. About two thirds of global DALYs are YLLs (and the rest YLDs), which means that the majority of disease burden is still due to premature death. Noncommunicable diseases are the leading cause of disease burden globally, and despite commonly being associated with wealth and a high standard of living, they are on the rise in low-income countries as well. The main results of the GBD 2023 study are reported in a series of articles [2, 28, 36]. A summary is provided in [37].

Mortality The mean age of death has increased between 2010 and 2023, and death rates have decreased in almost all age groups and regions. Thanks to vaccinations and prevention programs, this improvement has been especially prominent regarding many communicable diseases. The trends have not all been favorable, however. North America has recently seen increases in the death rates of children and young adults. Likewise, the death rates have seen dramatic local increases in Eastern Europe due to the Russian invasion of Ukraine, and in the Middle East due to the War in Gaza. [28]

The COVID-19 pandemic caused a temporary increase in deaths. In 2021, it was the leading cause of death globally. Now, mortality has returned to pre-pandemic levels in most countries and COVID-19 is no longer among the top global causes of death. [28] In some countries, however, the increase was larger and has persisted which shows the differences in healthcare systems, public policies, and socioeconomic status between countries [36]. The effects of the pandemic are more thoroughly analyzed in [33].

The most common causes of death globally in 2023 were heart disease, stroke, chronic obstructive pulmonary disease, lower respiratory infections, and neonatal disorders. Among the top 20 causes, the majority are NCDs, but there are still many treatable and even curable infectious diseases, e.g. diarrheal diseases on the 13th place and tuberculosis on the 15th place. [28]

Life expectancy Between 1950 and 2023, the global life expectancy has risen more than 25 years. It was 73.8 years globally in 2023 (for both sexes combined — women live longer than men almost everywhere). There are large differences in life expectancy between countries. In 2023, the life expectancy in Singapore was 85 years, while in some Sub-Saharan African countries it was under 60 years. [36]

Disease burden The total number of DALYs has grown globally from 2010 to 2023 but the age-adjusted rates have decreased by 13%. This is explained by population growth and aging. Importantly, the decrease in DALY rates shows that people are living longer and healthier lives. NCDs now account for the majority of disease burden which can be seen as a great success in the fight against CMNN diseases — DALY rates decreased by 40 to 50 percent for HIV, tuberculosis, Malaria, and diarrheal diseases in 2010–2023. The change from communicable to noncommunicable disease burden must be addressed in public policy, however. This can include promoting healthy food and physical activity, and ensuring equitable access to healthcare and treatments. [2]

The top causes of DALYs are similar to the top causes of death but diabetes, low back pain, and mental health disorders score noticeably higher due to their low mortality but high morbidity. The DALY rates from these health issues have also seen increases from 2010 to 2023. Anxiety disorders have increased 63%, depression 26%, and diabetes 13%. Among CMNNs, neonatal disorders and lower respiratory infections are the top causes of DALYs. [2]

The GBD study also quantifies the burden caused by specific risk factors. Half of global DALYs can be attributed to the 88 risk factors which are included in the analysis. The top five risk factors are high blood pressure, which alone accounts for 8.4% of all disease burden in the world, air pollution, high blood sugar, smoking, and low birth weight and short gestation. Globally, there have been 40–50% reductions in the disease burden attributable to some risk factors including unsafe sanitation, unsafe water sources, no access to handwashing, and child growth failure. Meanwhile, the burden from some causes has increased. Rises in high body mass index (BMI) and drug use have caused the largest increases. [2]

Future predictions According to GBD 2021, if current trends continue, global DALY rates are projected to decrease further from 2022 to 2050, although due to population growth and aging, total DALYs will increase. The reductions will be smaller than between 1990 and 2021, though. In addition, life expectancy will increase — most where it is the lowest now. Disease burden is projected to shift even further from CMNNs to NCDs, but this requires that funding for combatting CMNNs is not reduced. In addition, three alternative scenarios are analyzed by the GBD team: i) Safer Environment, ii) Improved Childhood Nutrition and Vaccination, and iii) Improved Behavioral and Metabolic Risks. The scenarios entail the gradual elimination of the respective risks from 2022 to 2050. The combined effect of these scenarios, if realized, is an additional 15% decrease in DALY rates. This shows that there are ways to substantially improve health outcomes beyond the current measures being taken. Different risk factors need to be eliminated in different countries; in Chad, for example, elimination of CMNN risks would yield a 10% reduction in DALYs, but in Germany the same reduction could be achieved by addressing NCD risk factors. [32]

3.2 Health interventions and their cost-effectiveness

Expenditure on healthcare and disease prevention can reduce disease burden — especially in the poorest countries

In wealthy countries, generally as much is spent on healthcare as is needed; if a treatment exists for a disease, sick individuals will get that treatment. In the poorest countries, however, the situation is very different. Many people lack access to the most basic healthcare, which is revealed, for example, in the high mortality rates of young children in many Sub-Saharan African countries — more than 10 times of that of high-income countries [36]. Extensive literature exists assessing the cost-effectiveness of specific health interventions, especially in low-income countries. There, foreign aid constitutes a large part of the very limited healthcare spending which is probably why CE is scrutinized. This subsection reviews the effectiveness of some health interventions including vaccinations, malaria prevention, sodium reduction for preventing heart disease, as well as smoking cessation campaigns.

Vaccines Both the costs and the effectiveness of vaccines can be reliably and straightforwardly calculated and thus many studies evaluate their CE. For example, COVID-19 vaccination is found to be very cost-effective or even cost-saving measure in mitigating the transmission of the virus and the effects of the pandemic [3]. Disease prevention with vaccines is generally very cost-effective, especially in low-income countries. A systematic review of vaccine CE literature in low- and middle-income countries finds a cost of under 100 USD per DALY averted in half of the studies [4]. Similarly, a study of simulated Malaria progression in Malawi finds vaccines to be a very cost-effective intervention (145 USD per DALY averted) [38].

Communicable diseases Among communicable diseases, the prevention and mitigation of Malaria, HIV, and diarrheal diseases are the most studied. The cost per DALY averted for Malaria interventions such as insecticide-treated nets is found to be as low as 50 USD and the cost of saving a life at just 2143 USD [5]. Averting a DALY with HIV prevention and control strategies varies from 199 USD to 799 USD depending on the intervention [6]. In [7], the CE of an integrated prevention campaign against Malaria, HIV, and diarrhea is assessed. The cost per DALY averted can be as low as 50 USD in Sub-Saharan Africa, for example, but in wealthier countries such as China and Mexico, the cost is over 10 000 USD.

Noncommunicable diseases Management of high blood pressure, a major risk factor for cardiovascular disease, is the focus of many health interventions studies. In a review of 42 hypertension management CE studies, interventions including medication and lifestyle advice are found to have costs in the range of hundreds of dollars per DALY averted in low- and middle-income countries. Generally, the wealthier the country, the less cost-effective the interventions are. [8] Similarly, another review of 42 diabetes and cardiovascular disease intervention studies in South Asia finds CE from cost-saving (negative price per DALY averted) to a few hundred USD per DALY averted for most interventions. This is well under the local annual GDP per capita. [9] For the prevention of hypertension, many sodium reduction strategies are also found to be cost-saving [10]. Finally, even small reductions in smoking prevalence can yield significant health benefits on a population level as tobacco is a major risk factor for many diseases, most notably cardiovascular disease and many cancers. A Mongolian study of tobacco control interventions finds that mass media campaigns can avert a DALY for just 24 USD, and that other interventions can be similarly very cost-effective [11].

Conclusions Consistently, the studies find very low prices (as low as under 100 USD) for many interventions for a saved year of healthy life, especially in the least wealthy countries. The costs are generally higher the wealthier the countries are. This is likely explained by the cheaper labor costs of poor countries and the fact that the cheapest health interventions are already applied in richer countries. The cost of a healthy year of life saved is generally much lower than the GDP per capita of the poorest countries. This suggests that, when used correctly, foreign aid and donations to non-governmental organizations can be very effective in reducing disease burden in some countries.

3.3 Similar work

In addition to specific interventions, literature also exists comparing the CE of all healthcare spending. In [39], the authors conduct a systematic literature review of all cost-per-DALY-averted studies and develop a registry which makes it easy to find them. 479 such studies are found from 2000 to 2015. Most of them are from Sub-Saharan Africa and focused on CMNN diseases. Primary prevention strategies (e.g. mosquito nets for malaria) are most commonly assessed, followed by vaccines and pharmaceutical interventions.

Most similar to our work is [12], where the authors quantify the association between healthcare spending and disease burden. They apply regression analysis to GDP, HDI, and disease burden data of different countries and find a nonlinear relationship between healthcare spending and age-adjusted DALY rate. In low HDI countries, one percent increase in healthcare expenditure is associated with an average 0.28% decrease in DALY rate. This corresponds to a cost of 998 USD per DALY averted. In very high HDI countries this associated cost is 69 499 USD. Compared with specific interventions such as vaccines [40] or mosquito nets [5], the estimated cost of a DALY averted (998 USD) is considerably higher with this purely associative methodology.

4 Methods

Our analysis includes four explanatory variables — GDP per capita (inflation- and purchasing power -adjusted), Gini coefficient, education index, and healthcare spending — and one response variable — age-adjusted DALY rate. Age-adjusted DALY rate has been chosen as the response variable because it is intuitive, widely used, comprehensive data is available, and it quantifies both ill health and early death. In addition, age-adjusted DALY rates enable comparisons between different countries and time periods. The socioeconomic indicators: Gini, GDP per capita, healthcare spending, and education index have been chosen due to their familiarity, and because they show the living conditions of countries from a wide perspective.

First, exploratory data analysis is performed: the data is analyzed with univariate statistics, then dependencies between the explanatory variables are examined using scatter plots and Spearman's correlation, and clusters are identified heuristically with principal component analysis and Mahalanobis depth. Multiple regression is used for modeling, that is, studying the relationship between the explanatory and response variables. The programming language `R` [41] is used throughout the work.

In the exploratory phase, healthcare spending is considered as a fraction of GDP — in relative terms. This allows for seeing differences between the countries more clearly. Healthcare spending is converted to dollars in the modeling phase, because the absolute monetary value is a more meaningful predictor of disease burden. Throughout the following sections, let us refer to healthcare spending as a fraction of GDP as "spending fraction" or "healthcare spending fraction", and healthcare spending in dollars as "spending" or "healthcare spending". Additionally, for conciseness, the age-adjusted DALY rate is sometimes referred to simply as "DALY" or "DALY rate".

4.1 Data sources

The data is obtained from IHME, IMF, UN, OECD, and World Bank data portals

Five datasets are used in our work. They contain the values of the variables by year and by country. They are obtained from the data portals by IHME [42], International Monetary Fund (IMF) [43], UN [44], OECD [45] and World Bank [14]. Table 4 shows a summary of the datasets and their sources.

Our analysis is limited to the years 1990–2023 due to the availability of the DALY data. Some countries have missing data points (years) in some datasets. Missing data points are filled in with linear interpolation, but some countries in some datasets are missing values from the start or at the end of the time period. These cannot be augmented with linear interpolation. Table 4 also shows the proportion of missing data before and after the augmentation.

The meaning and calculation of the variables in the DALY, Gini, and education datasets is discussed in Section 2. The GDP dataset contains inflation-adjusted purchasing power parity values. The health spending data shows the percentage of a country's GDP used for healthcare costs. [45] describes which costs are included in the calculations.

Dataset	**Data coverage**	**After augmentation**	**Source**
DALY rates	100 %	-	IHME [42]
Gini	71.9 %	86.5 %	World Bank [14]
Education index	100 %	-	UN [44]
GDP per capita	98.5 %	-	IMF [43]
Health spending	93.3 %	93.4 %	OECD [45]

Table 4: Summary of the datasets. All datasets contain data from all 38 OECD countries, except for Gini — New Zealand has no data. For some datasets and countries, the yearly data is partially incomplete. "Data coverage" shows the proportion of existing data points in 1990–2023 from the total (38 countries · 34 years = 1292 data points). "After augmentation" shows the coverage after linear interpolation ("-" indicates no change in data coverage).

4.2 Notation

Let n denote the number of observations and p the number of explanatory variables. In this thesis, $p = 4$ (GDP, education, Gini, and spending) and $n = 34 \cdot 38$ (number of years times number of countries). Scalars are denoted by lowercase italic letters, vectors by bold lowercase letters, and matrices by bold uppercase letters.

The explanatory data is collected in the matrix

$$\boldsymbol{X} = (x_{ij})_{n \times p},$$

where x_{ij} denotes the value of variable j for observation i, with $i = 1, \ldots, n$ and $j = 1, \ldots, p$.

The vector of explanatory variables for observation i is

$$\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top.$$

Here, $\boldsymbol{x}_i$ is denoted as a column vector and thus $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$.

The response observations are denoted by y_i, and the response vector by

$$\boldsymbol{y} = (y_1, \ldots, y_n)^\top.$$

For variable j, the sample mean and the standard deviation are

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \qquad \text{and} \qquad s_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2},$$

Similarly, for the response variable:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \qquad \text{and} \qquad s_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

The sample mean vector and covariance matrix of the explanatory variables are denoted by

$$\boldsymbol{\mu} = (\bar{x}_1, \ldots, \bar{x}_p)^\top, \qquad \text{and} \qquad \boldsymbol{\Sigma} = \frac{1}{n-1}\boldsymbol{X}^\top \boldsymbol{X}.$$

4.3 Univariate analysis

Univariate analysis utilizes basic visualizations and statistics

To obtain an overall picture of the data, univariate analysis is conducted. The mean, the median, and the coefficient of variation (also known as the relative standard deviation) are calculated for each of the five variables and analyzed numerically from tables and visually from charts, each by year and by country separately. For example, the average GDP per person is calculated by country, resulting in 34 values for the years 1990–2023 as well as by year, resulting in 38 values for the OECD countries. For variables where some countries have missing data (Gini, GDP, spending), the statistics are calculated for the time range when all countries have full data. Omitting this step would skew the results — if some country is missing GDP data from the beginning of the range 1990–2023, its average GDP would seem artificially high compared to other countries.

The coefficient of variation (CV) is calculated as the standard deviation divided by the mean. CV is used instead of standard deviation to show the relative changes in variation for data series that are increasing, and to enable comparison between variables with different units and scales. For example, GDP per capita has grown in most OECD countries and the standard deviation between countries has grown, but the relative differences in GDP per capita have not increased.

4.4 Correlation analysis

Scatter plots and Spearman's correlation help detect relationships between the variables

Based on initial results from the univariate analysis as well as for geographical comparison, we separate the countries into groups for the bivariate analysis. The chosen regions are: 1) the Americas with countries Canada, Chile, Colombia, Costa Rica, and Mexico; 2) East Asia and Oceania with countries Australia, Japan, New Zealand and South Korea; and 3) Europe with the rest of the countries (including Turkey and Israel). The United States is put into its own group (4) due to its dissimilarity with other countries in the Americas, especially in regard to healthcare spending.

Although the explanatory variables measure different quantities, we hypothesize them to be highly related. For example, as countries become richer (higher GDP per capita) they are likely to spend more on education, leading to increased education index (although the causation could be reverse as well). The dependencies between the explanatory variables are calculated using Spearman's correlation. In addition, the data is visually examined from scatter plots to detect nonlinear relationships. Missing data is excluded from the analysis. This includes Gini data for many countries, especially at the beginning of the time period, GDP data for some Eastern European countries in the beginning of the 1990s, and spending data for many countries at the end and at the beginning of the time period 1990–2023.

Spearman's correlation Spearman's correlation (or Spearman's ρ) measures monotonic dependence between variables. It varies between -1 and 1 and is similar to Pearson's correlation which measures linear dependence between variables.

Let $\boldsymbol{x}^{(j)} = (x_{1j}, \ldots, x_{nj})$ denote the vector of observations of variable j and $\boldsymbol{x}^{(k)} = (x_{1k}, \ldots, x_{nk})$ the vector of observations of variable k. Pearson's correlation is the normalized covariance between variables $\boldsymbol{x}^{(j)}$ and $\boldsymbol{x}^{(k)}$:

$$\rho_{\text{Pearson}}(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(k)}) = \frac{\text{Cov}(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(k)})}{s_j s_k},$$

where s_j and s_k are the standard deviations of variables j and k, respectively. Alternatively, we can also calculate the correlation between one of the explanatory variables and the response variable as follows:

$$\rho_{\text{Pearson}}(\boldsymbol{x}^{(j)}, \boldsymbol{y}) = \frac{\text{Cov}(\boldsymbol{x}^{(j)}, \boldsymbol{y})}{s_j s_y},$$

where s_y is the standard deviation of the response variable.

Spearman's correlation is calculated by ordering the variables and calculating Pearson's correlation between the rankings. Spearman's correlation differs from Pearson's correlation in that it yields larger values to nonlinear monotonic relationships.

The observations are converted to ranks by ordering the values of each variable separately. For variable j, let r_{ij} denote the rank of observation x_{ij} among $(x_{1j}, \ldots, x_{nj})$. The smallest value receives rank 1 and the largest value rank n. The same procedure is applied to variable k, yielding ranks r_{ik} for observations $(x_{1k}, \ldots, x_{nk})$.

After ranking, the original data pairs (x_{ij}, x_{ik}) are replaced by the rank pairs (r_{ij}, r_{ik}) for $i = 1, \ldots, n$. Spearman's correlation coefficient is then defined as the Pearson's correlation between the ranked variables:

$$\rho_{\text{Spearman}}(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(k)}) = \frac{\text{Cov}(\boldsymbol{r}^{(j)}, \boldsymbol{r}^{(k)})}{\text{SD}(\boldsymbol{r}^{(j)})\,\text{SD}(\boldsymbol{r}^{(k)})},$$

where $\boldsymbol{r}^{(j)} = (r_{1j}, \ldots, r_{nj})$ and $\boldsymbol{r}^{(k)} = (r_{1k}, \ldots, r_{nk})$ denote the rank vectors, and $\text{SD}(\cdot)$ denotes the standard deviation. This procedure can also be done such that the other of the two variables is the response variable $\boldsymbol{y}$.

In the case of a perfectly monotonically increasing relationship between the variables, the rankings coincide, i.e. $r_{ij} = r_{ik}$ for all i, which yields $\rho_{\text{Spearman}} = 1$.

We inspect Spearman's correlations instead of Pearson's due to the clear nonlinear relationships between some variables which can be identified in the scatter plots.

4.5 Heuristic clustering and outlier detection

Mahalanobis depth and principal component analysis are used to find outliers and clusters in the multivariate setting

A thorough understanding of the data requires recognizing which countries differ significantly from the others. This information can also be used in multiple regression to separately analyze similar groups of countries (clusters). Heuristic clustering is conducted using principal component analysis and Mahalanobis depth. The methods are robustified by using the minimum covariance determinant (MCD) method to estimate the location and scatter functionals (mean vector and covariance matrix). Outlier detection is conducted separately on the explanatory variables and the response variable. For the response variable, the DALY rate, only Mahalanobis depth is used. In addition, outliers are detected visually from time series line plots.

4.5.1 Principal component analysis

Principal component analysis (PCA) is a multivariate analysis technique for dimensionality reduction. It finds linear combinations of the variables (called *principal components*) such that they are uncorrelated with each other and explain as much of the variance in the data as possible. The first component is chosen such that it explains the maximum amount of variance, and the second component such that it is orthogonal to the first and explains as much of the remaining variance as possible. The adding of components is repeated until the principal components explain a desired proportion of variation in the data. Usually PCA is used for dimensionality reduction, but it can be also used for our purpose — heuristic clustering and outlier detection — by visualizing the principal component scores. The following gives a concise mathematical description of sample PCA. More details can be found in [46].

As introduced in Section 4.2, the data matrix is

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \in \mathbb{R}^{n \times p},$$

where n is the number of observations and p is the number of variables.

The variables are centered and standardized to unit variance because they have different scales:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j},$$

where $\bar{x}_j$ is the sample mean and s_j the sample standard deviation of variable j.

The sample correlation matrix is obtained by:

$$\boldsymbol{C} = \frac{1}{n-1} \boldsymbol{Z}^\top \boldsymbol{Z} \in \mathbb{R}^{p \times p},$$

where $\boldsymbol{Z} = (z_{ij})_{n \times p}$ is the centered and standardized data matrix.

PCA is based on the eigenvalue decomposition of the correlation matrix:

$$\boldsymbol{C}\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k, \qquad k = 1, 2, \ldots, p,$$

where λ_k are the eigenvalues which correspond to the proportion of the variance in the data explained by principal component k, and $\boldsymbol{v}_k$ are the corresponding eigenvectors (principal components).

The eigenvalues are indexed in decreasing order:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0.$$

The ordered eigenvectors form the matrix of loading vectors:

$$\boldsymbol{L} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k] \in \mathbb{R}^{p \times k},$$

where $k \leq p$ is the number of principal components chosen. The loadings show how much the original variables contribute to each of the principal axes. Thus, they give a way to interpret the meanings of the principal axes. For example, if GDP per capita and education index have high loading scores for the first principal component $\mathbf{v}_1$, this means that together they explain the most variation in the data, and that the first principal component is a "wealth and education axis".

The principal component scores are obtained by projecting the data to the space spanned by the principal axes:

$$\boldsymbol{T} = \boldsymbol{Z}\boldsymbol{L} \in \mathbb{R}^{n \times k}.$$

The rows of $\boldsymbol{T}$ are the coordinates of the observations along the principal axes. Using the scores, the observations can be plotted in two ($k = 2$) or three ($k = 3$) dimensions. We utilize this approach in visually identifying outliers and clusters among the countries.

The proportion of variance explained by the k-th principal component is given by

$$\text{Explained Variance}(k) = \frac{\lambda_k}{\sum_{j=1}^{p} \lambda_j}.$$

The cumulative explained variance by the first k principal components is

$$\text{Cumulative Variance}(k) = \frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{p} \lambda_j}.$$

If the first 2 or 3 components explain a large proportion of the variance in the data, e.g. 80% or 90%, the plots can be used reasonably well to make inferences about the data.

4.5.2 Mahalanobis depth

Another way to detect outliers is by using statistical depth. Depth functions measure the centrality of an observation $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ with respect to a set of p-variate observations in $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$. The value of depth functions varies from 0 to 1 with higher values indicating more centrality with respect to the observations in $\boldsymbol{X}$. Being one of the simplest and most widely known depth functions, we use Mahalanobis depth to identify atypical countries with respect to the explanatory variables or to the response variable. For more details on Mahalanobis depth, see [47].

Mahalanobis depth is defined as follows:

$$D_M(\boldsymbol{x}_i) = \frac{1}{1 + d_M(\boldsymbol{x}_i)^2},$$

where

$$d_M(\boldsymbol{x}_i) = \sqrt{(\boldsymbol{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu})},$$

and $\boldsymbol{\mu}$ is the sample mean and $\boldsymbol{\Sigma}$ the sample covariance matrix.

The Mahalanobis depths are calculated for each country by using the averages for each variable in the years 2015–2019. All countries have data for all variables in this time period, except for New Zealand, which is why it is excluded from the outlier analysis.

4.5.3 Robust scatter functionals

Due to the presence of outliers which can be identified in the univariate and bivariate analyses, the principal component analysis and the Mahalanobis depth calculations are repeated with the minimum covariance determinant (MCD) method [48]. Using this method, a proportion α of the data is sampled and the mean vector and the covariance matrix are calculated from the sample. This sampling is repeated multiple times and the final mean vector and the covariance matrix are chosen from the sample with the smallest covariance matrix determinant. A value of $\alpha = 0.75$ is used for PCA and $\alpha = 0.5$ for the Mahalanobis depth. When PCA is no longer performed using the covariance matrix of the full dataset, it should be more accurately called a robust PCA-type transformation, but we shall refer to it simply as robust PCA.

4.6 Multiple regression

Linear regression with variable transformations is used

To observe the effects of all explanatory variables together on the response variable, and to examine the relationship more quantitatively, regression analysis is used. In regression analysis, the response variable (DALY rate) is predicted from the explanatory variables such that they explain as much of its variation as possible. Let us refer to age-adjusted DALY rate as "DALY", GDP per capita as "GDP", Gini coefficient as "Gini", healthcare spending fraction as "SF", and education index as "EI". Only rows with complete data (values of all the variables for the country-year-pair) are used in the analysis. This reduces the number of available observations from 1292 to 1032. New Zealand is not present in the analysis due to the complete lack of Gini data.

The regression analysis is conducted using linear regression with variable transformations. Linear regression provides models that are both interpretable and that quantify the statistical significance and the magnitude of relationships, which is why it is ideal for answering our research question. The linear regression models are fitted to the whole data as well as individually on the clusters identified in PCA. The predictive power of the explanatory variables is tested individually as well as together. The regression analysis is conducted on the dataset such that the years of the COVID-19 pandemic (2020–2023) are removed. The pandemic represents a shock to disease burden which is not well-captured by the explanatory variables.

The following sections give a brief introduction to the topics relevant to our analysis including multiple linear regression, heteroscedasticity, coefficient of determination, variable transformations, and adjusting for covariates as well as significance testing. An introduction to regression modeling can be found in [49]. More advanced methods specifically applied to panel data can be found in [50].

4.6.1 Clustering

There are three clusters of countries (more accurately, country-year-pairs) that are identified in the principal component analysis. The clusters consist of countries that have similar combinations of the values of the explanatory variables. The clustering might be caused by political, cultural, historical, or other factors which are not captured in the explanatory variables. The linear regression analysis is conducted on the clusters separately in addition to the data as a whole. This allows for interpreting the effects of the explanatory variables on the DALY rate within similar countries.

4.6.2 Multiple linear regression

Multiple linear regression predicts the value of a response variable from a set of explanatory variables. As indicated by its name, the response variable is assumed to be linearly dependent on the explanatory variables. We add the year as one of the predictors. Thus, the number of variables becomes $\tilde{p} = p + 1$ in the regression analysis.

Let $\boldsymbol{y} = (y_1, \dots, y_n)^\top$ denote the response vector, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{\tilde{p}})^\top$ the regression coefficient vector, $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ the residuals, and $\tilde{\boldsymbol{X}} \in \mathbb{R}^{n \times \tilde{p}}$ the *design matrix*, which is defined as:

$$\tilde{\boldsymbol{X}} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} & x_{1\tilde{p}} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} & x_{2\tilde{p}} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} & x_{n\tilde{p}} \end{pmatrix}.$$

The multiple linear regression model is given by:

$$\boldsymbol{y} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

The residuals are assumed to be uncorrelated, have zero mean ($\mathbb{E}[\boldsymbol{\varepsilon}] = \boldsymbol{0}$) and equal variance ($\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n$, where σ^2 denotes the noise variance and $\boldsymbol{I}_n$ is the $n \times n$ identity matrix).

The regression coefficients are estimated by minimizing the sum of squared residuals:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\|_2^2.$$

The resulting ordinary least squares (OLS) estimator is:

$$\hat{\boldsymbol{\beta}} = (\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^\top \boldsymbol{y},$$

when $\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}}$ is invertible.

The fitted values are given by $\hat{\boldsymbol{y}} = \tilde{\boldsymbol{X}} \hat{\boldsymbol{\beta}}$, and the residual vector by $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \hat{\boldsymbol{y}}$.

Each coefficient β_j represents the marginal effect of the j-th explanatory variable on the response variable, holding all other variables constant.

4.6.3 Significance testing

The significance of the predictors can be tested by estimating the standard errors and conducting a t-test as follows. The significance test described here assumes that the residuals are normally distributed. The standard errors are calculated as:

$$\mathrm{SE}(\hat{\beta}_j) = \sqrt{[\mathrm{Var}(\hat{\boldsymbol{\beta}})]_{jj}},$$

where $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 (\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}})^{-1}$. $\hat{\sigma}^2$ is the estimated variance of the residuals.

The test statistic is the following:

$$t_j = \frac{\hat{\beta}_j}{\mathrm{SE}(\hat{\beta}_j)}.$$

The null hypothesis is that the explanatory variable does not have an effect on the response variable (or that the data is insufficient to conclude it):

$$H_0 : \beta_j = 0.$$

Under the null hypothesis, the test statistic follows the t-distribution $t_j \sim \mathtt{t}_{n-\tilde{p}-1}$ and thus the p-value for coefficient j is calculated as:

$$\text{p-value}_j = \mathbb{P}(|t_j| \leq |\mathtt{t}_{n-\tilde{p}-1}|).$$

4.6.4 Heteroscedasticity

In standard linear regression it is assumed that the residuals ε_i are homoscedastic, that is, their variance does not depend on the values of the variables. However, in our data this is not the case. As can be observed in the univariate analysis, the variance increases as the variables have larger values. We confirm the presence of heteroscedasticity with the White test and the Breusch-Pagan test. Ordinary least squares gives an unbiased estimate of the coefficients, but, in the presence of heteroscedasticity, the p-values are not accurate. We use a heteroscedasticity-consistent variance estimator [51] to obtain accurate estimates for the p-values.

In a model with heteroscedasticity:

$$\mathrm{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Omega},$$

where $\boldsymbol{\Omega}$ is diagonal, but the entries are not constant. Thus,

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = (\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^\top \boldsymbol{\Omega} \tilde{\boldsymbol{X}} (\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}})^{-1}.$$

To estimate the p-values correctly, we use White's "sandwich" estimator:

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = (\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^\top \widehat{\boldsymbol{\Omega}} \tilde{\boldsymbol{X}} (\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}})^{-1},$$

where weighted squared residuals are used for the estimate of $\widehat{\boldsymbol{\Omega}}$, i.e.,

$$\widehat{\Omega}_{ii} = \omega_i \hat{\varepsilon}_i^2,$$

where the weights

$$\omega_i = \frac{1}{(1 - h_{ii})^2}$$

are calculated using the leverage $h_{ii} = \boldsymbol{x}_i^\top (\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}})^{-1} \boldsymbol{x}_i$ which measures how influential observation i is. This makes the method robust to outliers and the variance is not underestimated.

The adjusted p-values are obtained by applying the t-test to the revised $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})$ as described previously.

4.6.5 Coefficient of determination

Coefficient of determination ("R squared" or R^2) is a measure of the predictive ability of a regression model. It varies between 0 and 1 with $R^2 = 1$ indicating that the model perfectly explains the variance in the data and $R^2 = 0$ indicating a useless model. More specifically, the value of R^2 is the proportion of variation in the response variable explained by the explanatory variables. R squared is calculated using the sum of squared residuals (SS_{res}) and the total sum of squares (SS_{tot}) as follows:

$$R^2 = 1 - \frac{\mathrm{SS}_{res}}{\mathrm{SS}_{tot}},$$

where

$$\mathrm{SS}_{res} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \quad \text{and} \quad \mathrm{SS}_{tot} = \sum_{i=1}^{n} (y_i - \bar{y})^2,$$

and $\bar{y}$ is the mean of the response variable.

Notably, R^2 can only increase as more predictors are added to the model — even if they are completely unrelated to the response variable. This is because completely uncorrelated predictors will get a coefficient β_j of zero, and if by change they are correlated with the response variable, they "improve" the model. Adjustment to R^2 tries to account for this by penalizing the number of variables. We use the adjustment proposed in [52]. Let us refer to it as the adjusted R^2 (R^2_{adj}). It is calculated as follows:

$$R^2_{adj} = 1 - \frac{\mathrm{SS}_{res}/\mathrm{df}_{res}}{\mathrm{SS}_{tot}/\mathrm{df}_{tot}},$$

where df_{res} and df_{tot} denote the degrees of freedom of the population variance around the model and the mean, respectively. Thus,

$$R^2_{adj} = 1 - (1 - R^2) \frac{n - 1}{n - \tilde{p} - 1}.$$

4.6.6 Variable transformations

Nonlinear dependencies can be modeled with linear regression by transforming the variables. For example, in the bivariate analysis it can be observed that GDP seems to have a logarithmic relationship with DALY rate. This observation supports the intuitive notion of diminishing returns from additional wealth. In addition, we transform the spending fraction variable because it is not a meaningful predictor of disease burden, whereas the absolute spending in dollars is.

We apply the following two transformations:

1. SF $\rightarrow \log(\mathrm{SF} \cdot \mathrm{GDP}) = \log(\mathrm{SP})$, where "SP" refers to healthcare spending.

2. GDP $\rightarrow \log(\mathrm{GDP})$. To better represent the nonlinear relationship.

Both transformations are tested individually to see that they increase R^2_{adj}, i.e., both improve the regression model.

Finally, the year is added as one of the explanatory variables to the model. It is included to capture the effect of increased knowledge in the medical field as well as changes in lifestyle, e.g., increased awareness of the dangers of smoking. Thus, the final regression model is:

$$\mathrm{DALY} = \beta_0 + \beta_1 \cdot \log(\mathrm{GDP}) + \beta_2 \cdot \mathrm{EI} + \beta_3 \cdot \mathrm{Gini} + \beta_4 \cdot \log(\mathrm{SP}) + \beta_5 \cdot \mathrm{year} + \varepsilon \quad (1)$$

4.6.7 Univariate regression models

As can be observed in the bivariate analysis, there is strong collinearity among the explanatory variables. For example, as countries become richer (higher GDP), they tend to become more educated (higher EI). Because of this, it is hard to interpret the relative importance of the different factors from the full regression model. Therefore, the variables are tested individually. As an example, the strength of education index as an individual predictor can be tested with the following model:

$$\mathrm{DALY} = \beta_0 + \beta_1 \cdot \mathrm{EI} + \varepsilon.$$

Explanatory variables which produce a univariate regression model which achieves the highest coefficient of determination are the strongest predictors of disease burden.

4.6.8 Adjusting for covariates

In addition to comparing the R^2 scores of the univariate models, the effect of explanatory variable j on the response variable can be isolated by comparing its associated coefficient in the full model with all variables ($\beta_{j,full}$) to its coefficient in the univariate model ($\beta_{j,univ.}$). We conduct this analysis for the logarithmic spending $\log(\mathrm{SP})$ to more carefully examine its effects.

5 Results

In this section, we present the results from the univariate, bivariate, and multivariate analyses. Key findings are summarized under each subsection heading. The univariate results compare the averages between countries and show historical developments in the five variables. The most interesting findings are visualized with maps and line graphs. Bivariate relationships are shown with scatter plots and correlation coefficient tables. Heuristic clustering using principal component analysis and statistical depth reveal three distinct clusters of countries. The PCA results are presented with biplots and the depth function analysis with bar charts. Multiple regression results conclude the section. Different models are compared using their adjusted R^2 scores, and the significance of predictors is reported with p-values. The most important predictor, logarithmic healthcare spending, is analyzed more carefully.

5.1 Univariate analysis

Age-adjusted DALY rates, GDP per capita, and education have improved between 1990 and 2023 but large differences between countries remain

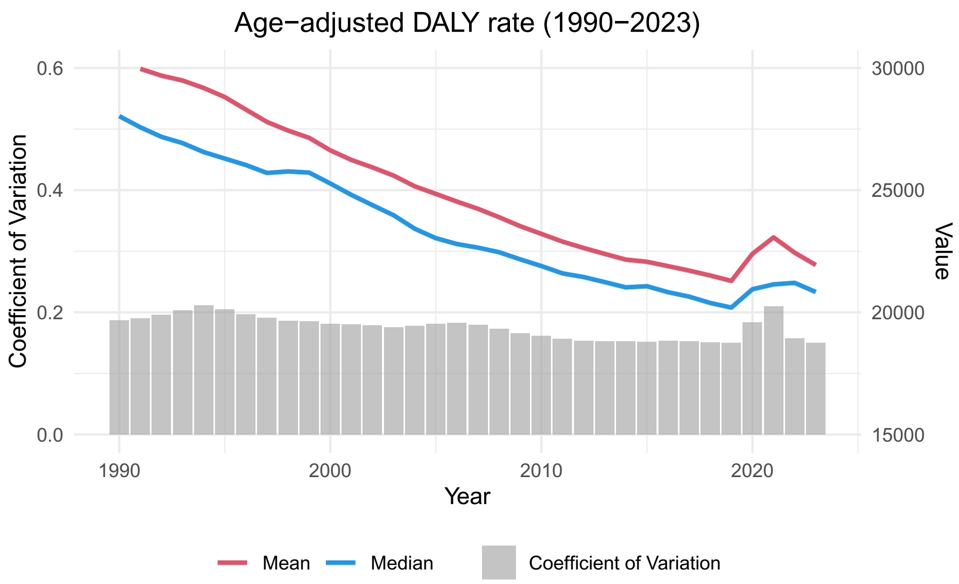

Figure 4: Mean and median age-adjusted DALY rate from 1990 to 2023, as well as the coefficient of variation.

DALY rates The mean age-adjusted DALY rate has decreased from approximately 30 000 to 22 000 per 100 000 from 1990 to 2019, meaning that, on average, people are living longer and being healthier in the same age groups (see Figure 4). The years of the COVID-19 pandemic (2020–2023) can be seen as outliers with a maximum increase to 23 000 in 2021. The median has been consistently below the mean, which suggests that disease burden is unevenly distributed between countries. The coefficient of variation has also slightly decreased from 0.19 in 1990 to 0.15 in 2025, meaning that the differences between OECD countries have shrunk. The years 1993–1995 and 2005–2008 show a slight increase in the coefficient of variation of 0.01. In addition, the years 2020 and 2021 show a large increase to 0.18 and 0.21 from the 0.15 in 2019, which suggests that the pandemic affected countries unequally.

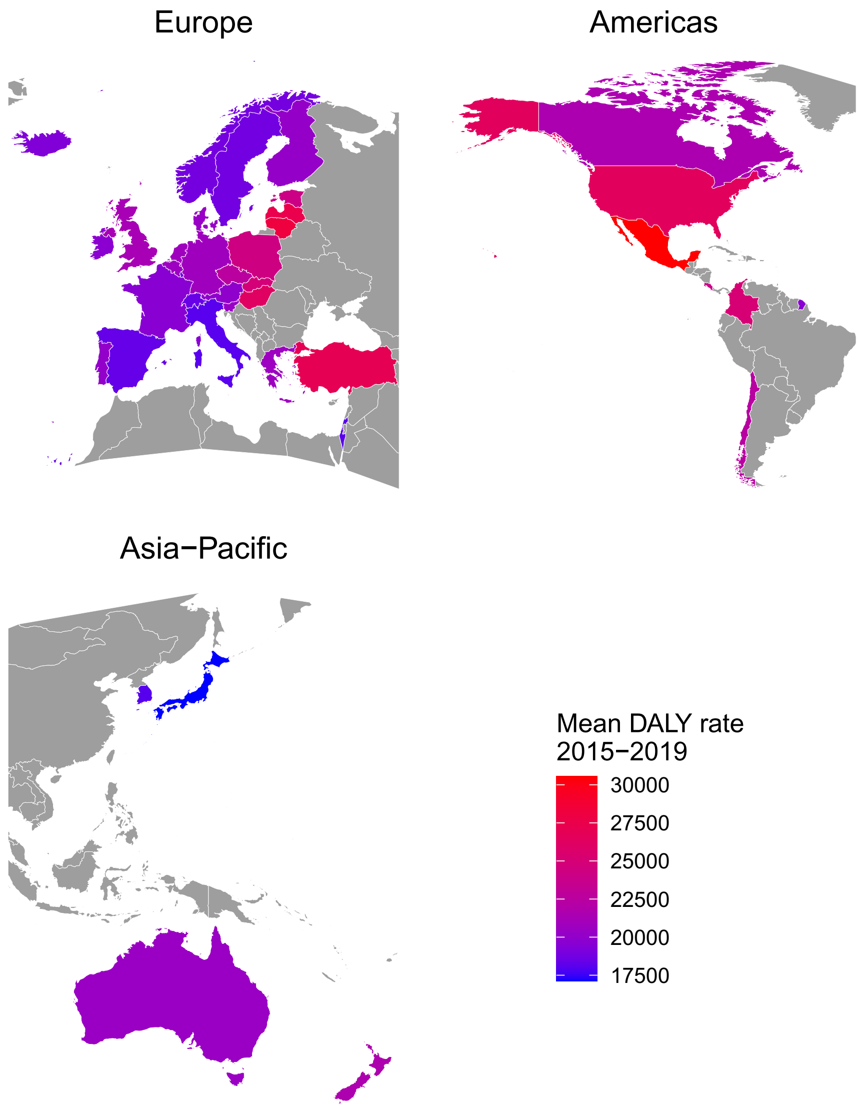

Figure 5: The average age-adjusted DALY rate between 2015 and 2019 by country. To improve visibility of small nations, the world map is broken into three enlarged parts.

When examining by country, large differences can be found. A heatmap of mean age-adjusted DALY rates between 2015–2019 is presented in Figure 5. The year range is chosen to reflect the current situation without the effects of the COVID-19 pandemic. Five countries with the highest disease burden are Mexico, Lithuania, Latvia, Turkey, and the US with rates between 26 300 and 30 600 per 100 000. Meanwhile, the lowest disease burden is seen in Japan, Korea, Italy, Israel, and Spain, varying from 17 100 to 18 400. The difference between the lowest average rate (Japan, 17 100) and the highest (Mexico, 30 600) is considerable. To put it into perspective, Japan would achieve approximately the same rate as Mexico if every one of its citizens suffered from uncontrolled asthma (see Table 2) for their entire lives.

GDP The mean GDP per capita in OECD countries has increased from 18 000 USD in 1995 to over 60 000 USD in 2023 as can be seen from Figure 6. The year range has been determined by missing data from some Eastern European countries from the beginning of the 1990s. The median GDP per capita has been above the mean until 2004, after which there has been a reversal and the difference is growing. This implies that some countries have become considerably richer than others. A closer inspection reveals that the reversal might be explained by tax havens such as Ireland, Luxembourg, and Switzerland, becoming very rich. The coefficient of variation has been considerably higher (0.4–0.5) than that of the DALY rate (0.15–0.20), meaning that the relative differences between countries are larger. The financial crisis of 2008 and the start of the COVID-19 pandemic can be seen as small temporary decreases in the GDP per capita.

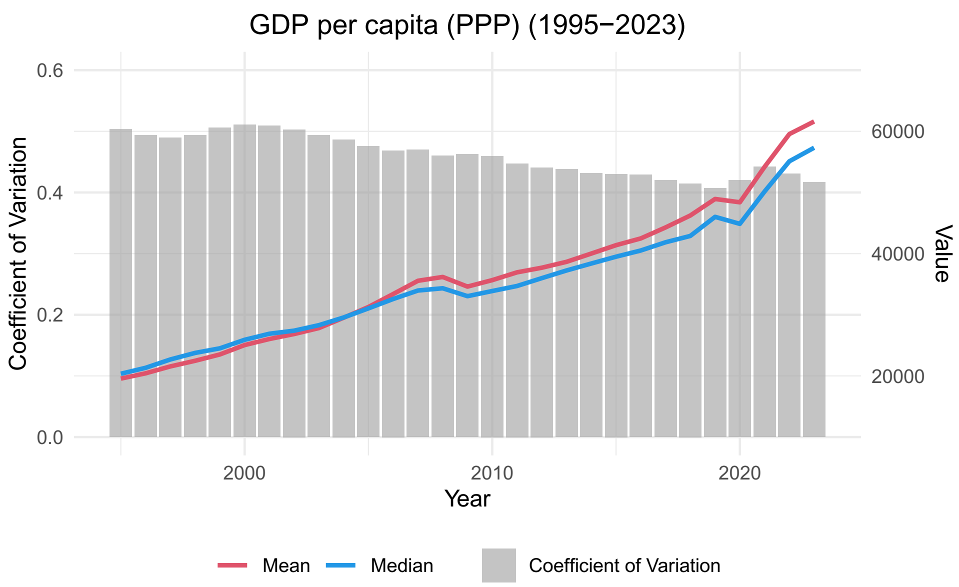

Figure 6: Mean and median GDP per capita from 1995 to 2023, as well as the coefficient of variation.

Similarly to the DALY rate, large differences between countries can be found in GDP per capita (as suggested by the large coefficient of variation). Again, the averages from the recent years 2015–2019 are examined. Colombia has a GDP per capita of 14

974 USD whereas Luxembourg boasts a GDP of 116 657 USD. Other outliers are Ireland (80 417 USD), Switzerland (69 269 USD), and Norway (65 807 USD). For other countries, GDP per capita ranges from 20 000 USD to 61 000 USD. The lowest levels are found in Central and South America and Eastern Europe, and the highest values in Southern and Western Europe.

Education The average education index has steadily increased from 0.70 in 1990 to 0.85 in 2023, although the rate of increase has been decreasing as can be seen in Figure 7. During the same period, the variation between countries has decreased as indicated by the coefficient of variation going from 0.16 to 0.08. Turkey has the largest change (going from 0.4 to 0.8) while the United States has the smallest (from 0.87 to 0.91). In the year 2023, Iceland (0.96), Germany (0.96), and the UK (0.94) had the highest level of education and Colombia (0.7), Mexico (0.71), and Costa Rica (0.75) the lowest. Other countries had values between 0.8 and 0.94. The education index is quite high in all countries, meaning that the populations of OECD countries are on average highly educated — even the lowest value (0.7) would be achieved only by, for example, 11 years of mean schooling and 12 years of expected schooling.

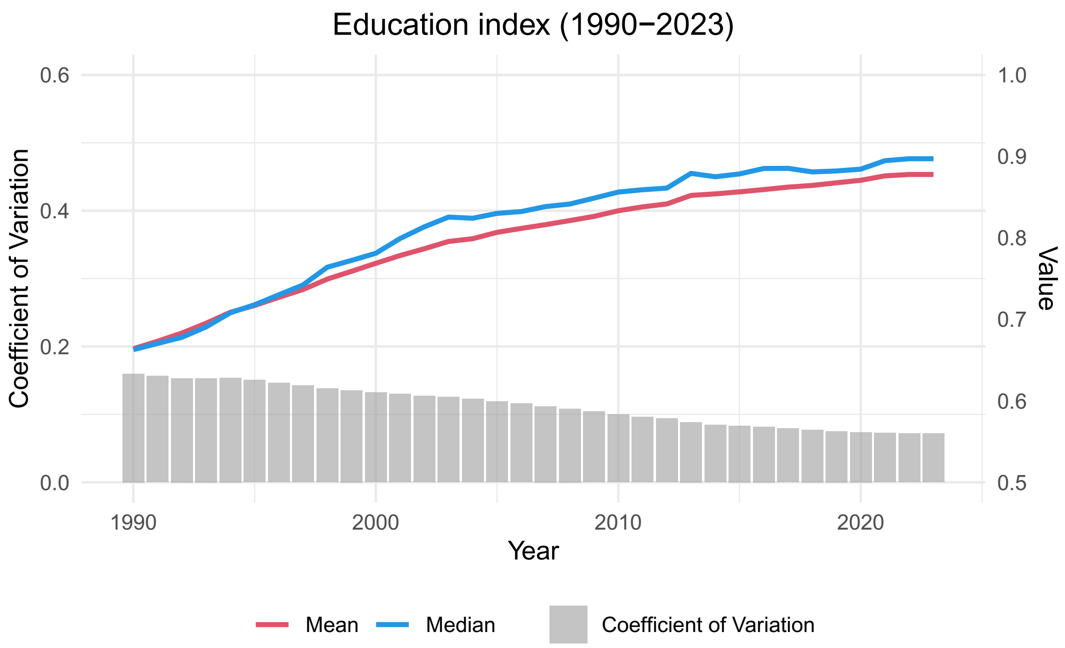

Figure 7: Mean and median, and the coefficient of variation of education index from 1990 to 2023.

Healthcare spending fraction Similarly to most other variables, the mean healthcare spending as percentage of GDP shows growth from 7% in 2004 to almost 9% in 2023 (Figure 8). It should be noted that since the GDP per capita and the populations of most countries have grown, the total healthcare spending (GDP times spending as fraction of GDP) has grown faster than linearly. Again, the year range has been chosen such that all countries have data within the time period. The years 2020–2021 show an increase to almost 10% due to the COVID-19 pandemic. The yearly CV between countries has stayed around 0.25, meaning that the relative differences between countries have remained the same. Large differences in spending fraction exist between countries. When looking at the recent averages 2015–2019, the highest average spender, the United States, uses almost 17% of its GDP for healthcare, whereas Turkey spends less than 5%. Other countries fall between 5% and 11%.

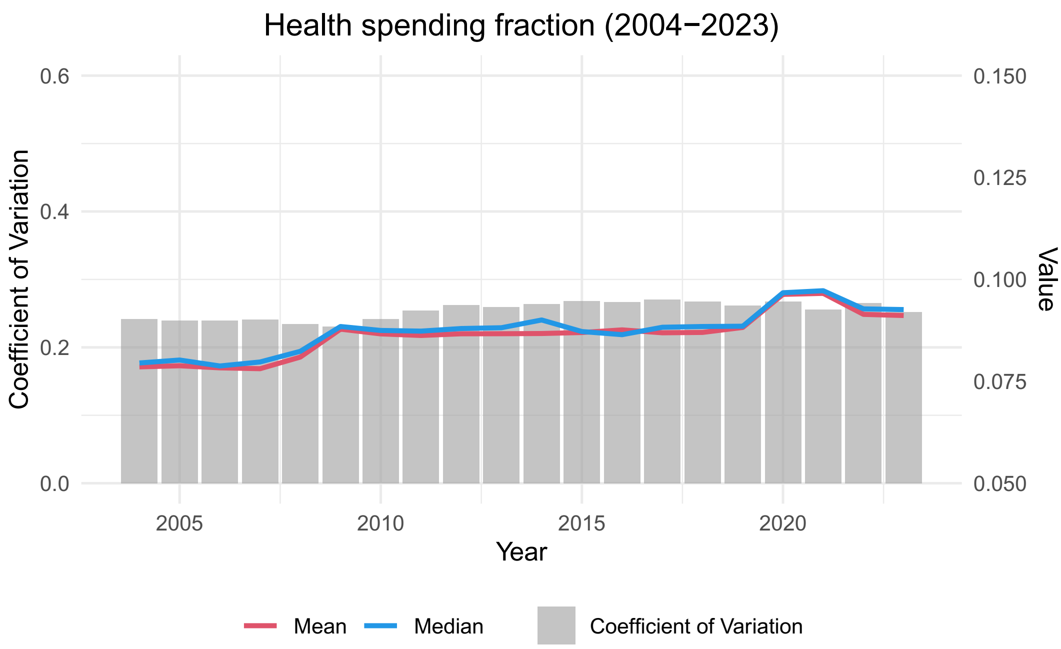

Figure 8: Mean and median, and the coefficient of variation of healthcare spending fraction from 2004 to 2023.

Gini The mean Gini has stayed at a constant level of 0.35 throughout 2008 to 2017. These are the only years for which every country has data (except for New Zealand which has no Gini data). The mean and the median have been very similar and the coefficient of variation has stayed around 0.2. Figure 9 shows the average Gini in 2015–2017. The OECD countries in the Americas have the highest mean Gini coefficients. Colombia, Costa Rica, Mexico, and Chile have values between 0.45 and 0.50. Turkey, the US, and Israel have values that are very close (0.39–0.43). Central European and Nordic countries generally have the lowest values of the Gini coefficient ranging from 0.25 to 0.3.

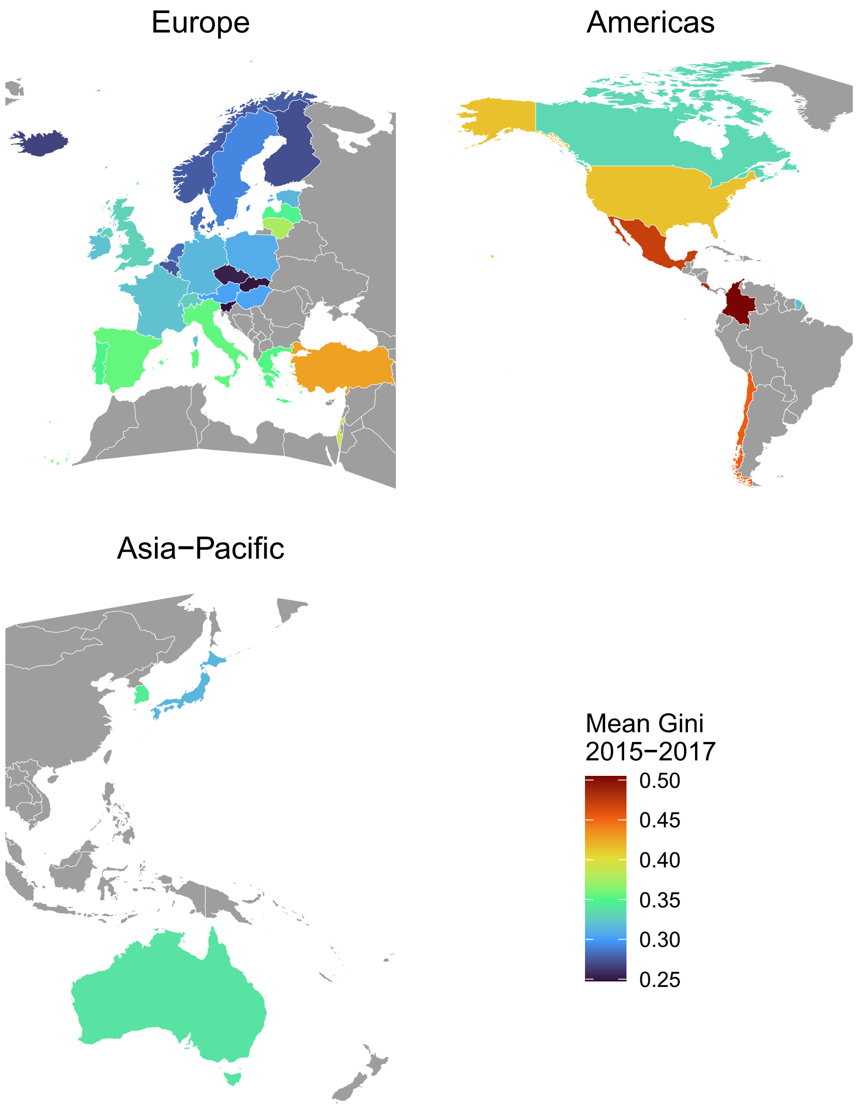

Figure 9: The average Gini from 2015 to 2017.

5.2 Correlation analysis

Higher wealth, education and health spending, and lower inequality seem to be associated with lower disease burden

As hypothesized, there are clear relationships between the explanatory variables as well as with the response variable. Lower DALY rates are associated with a higher GDP per capita, higher education index, and higher spending fraction. Higher Gini index is linked to higher disease burden, especially when comparing between regions. Countries with higher GDP tend to be more economically equal, more educated, and spend more on healthcare (even as a fraction of GDP). Similarly to Section 4.6, let us refer to age-adjusted DALY rate as "DALY", GDP per capita as "GDP", Gini coefficient as "Gini", healthcare spending fraction as "SF", and education index as "EI".

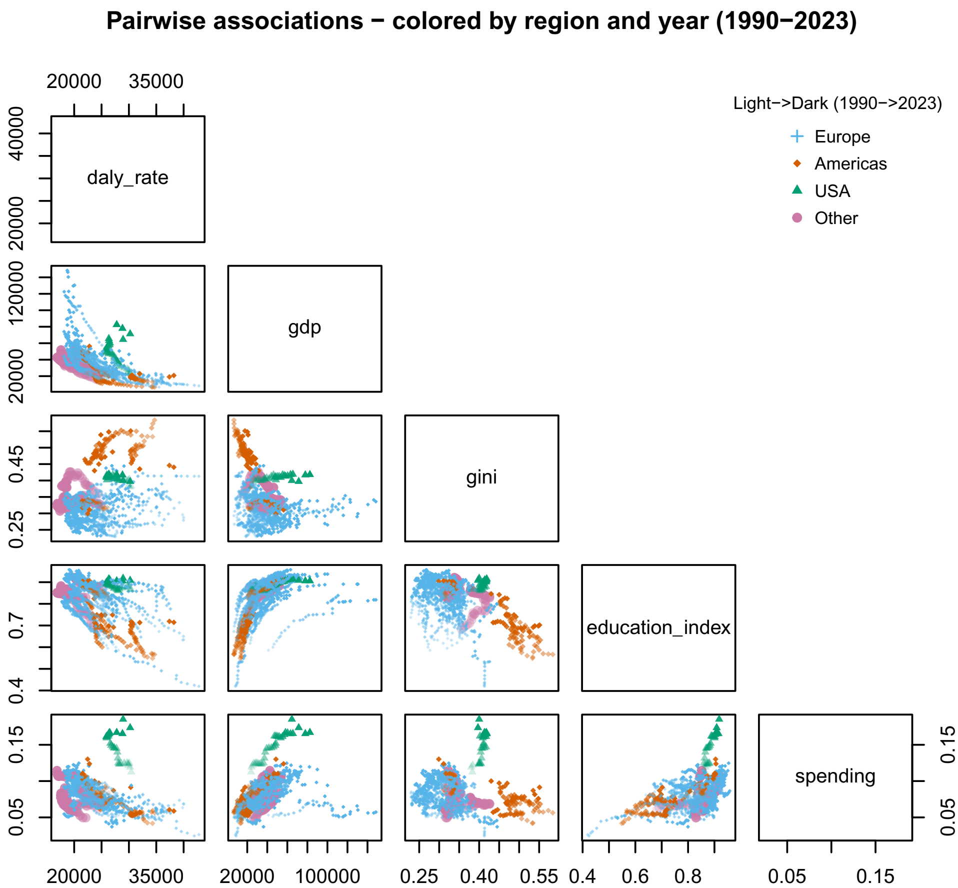

Figure 10: Pairwise scatter plots of the variables. "spending" denotes spending fraction in the graph.

Figure 10 shows pairwise scatter plots between all variables. The data points are the year-country -pairs. In many of the scatter plots, the geographical groups

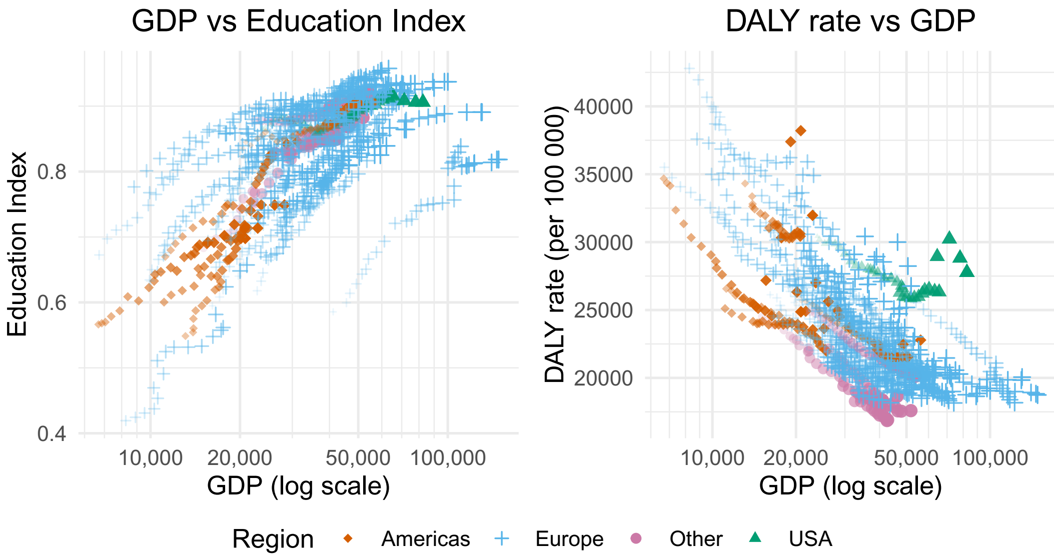

Figure 11: Logarithmic plots of the DALY–GDP and the GDP–EI -relationships. Higher opacity and larger size of the marker indicate more recent data points.

are clearly separated. Gini separates European countries from American countries except for Canada. In healthcare spending fraction, the United States is different from other countries. Some relationships, such as SF–GDP, EI–GDP and EI–DALY appear approximately linear while others such as GDP–DALY and GDP–EI show a monotonic but nonlinear relationship. For example, there seems to be a positive but a diminishing relationship between GDP and EI. The most scattered is the plot between Gini and DALY, meaning that multivariate analysis is required to fully understand their relationship.

The relationship between DALY and GDP as well as between GDP and EI are more closely examined in Figure 11. In the figure, the scatter plots are represented such that the GDP axis has logarithmic scale. This transformation makes the GDP–DALY -relationship approximately linear and the GDP–EI -relationship almost linear. The imperfectly logarithmic GDP–EI -relationship is likely due to the way that the education index is calculated — if the expected schooling is over 18 years or the average schooling over 15 years, the index does not increase. Despite the better fit of the logarithmic relationship between DALY and GDP, there are still large differences between countries. Some countries such as the United States seems to be getting "much less health at a given level of wealth". On the other hand, Japan and Korea are the least sick despite their unexceptional GDP per capita.

Table 5 shows the Spearman's correlations between the variables as well as the breakdown by region. The correlations are in accordance with the results from the visual analysis. Especially strong associations are DALY–GDP, GDP–EI, GDP–SF, and SF–DALY. The regional breakdown shows that there are large differences between the correlations between regions. The United States, especially, often has reverse correlations compared to the overall correlations. This is the case with DALY–Gini, GDP–Gini, Gini–EI, and Gini–SF. While in other regions, economic inequality is associated with a lower GDP, in the US the historical development is the opposite.

Table 5: Spearman correlations by region.

Pair	Overall	Europe	USA	Americas	Other
DALY — GDP	-0.684	-0.758	-0.561	-0.743	-0.641
DALY — Gini	0.236	0.131	-0.710	0.626	0.322
DALY — EI	-0.331	-0.330	-0.420	-0.763	-0.132
DALY — SF	-0.477	-0.632	-0.528	-0.795	-0.513
GDP — Gini	-0.384	-0.296	0.697	-0.843	-0.337
GDP — EI	0.617	0.522	0.841	0.944	0.747
GDP — SF	0.542	0.521	0.960	0.761	0.589
Gini — EI	-0.427	-0.391	0.548	-0.850	-0.355
Gini — SF	-0.196	-0.239	0.629	-0.669	-0.434
EI — SF	0.467	0.399	0.871	0.848	0.399

5.3 Heuristic clustering and outlier detection

5.3.1 Explanatory variables

The explanatory variables divide the countries into three clusters: the US, Central and South America and Turkey, and the rest

PCA Principal component analysis suggests that the countries form three approximate clusters when examining the four explanatory variables. The clusters are the same both with regular and robust PCA (Figure 12). The clusters are similar to the geographical groups (Europe, the Americas, other) examined in Section 5.2 but with some important differences. Canada, Israel, Australia, Japan, and South Korea seem similar to the European countries, while Turkey is dissimilar from other European countries. Although the most recent data points from Turkey are close to the European and Asian countries, the time series of the country as a whole is closer to the Central and South American countries. The data points from the United States can be seen as clearly separate from other countries. Thus, the following clustering improves upon the geographical grouping: 1) Europe, Asia, Oceania, and Canada, 2) Central and South America and Turkey, and 3) the US. Let us refer to these clusters as: 1) high-income, 2) middle-income, and 3) the US. We use these new three clusters in the regression analysis. The new clustering is visualized in Figure 13 with new colors.

The loading scores are shown in Figure 12 using arrows. They represent the contribution of each variable on the construction of the principal axes. Close alignment with an axis and a long length signifies high contribution to the axis. The loading scores show that Gini separates the US, Turkey, and Central and South America from the rest, and that healthcare spending separates the US from the other countries with higher inequality. In regular PCA, the first two principal components explain approximately 77% of the variance, and using robust PCA, about 80% of the variation

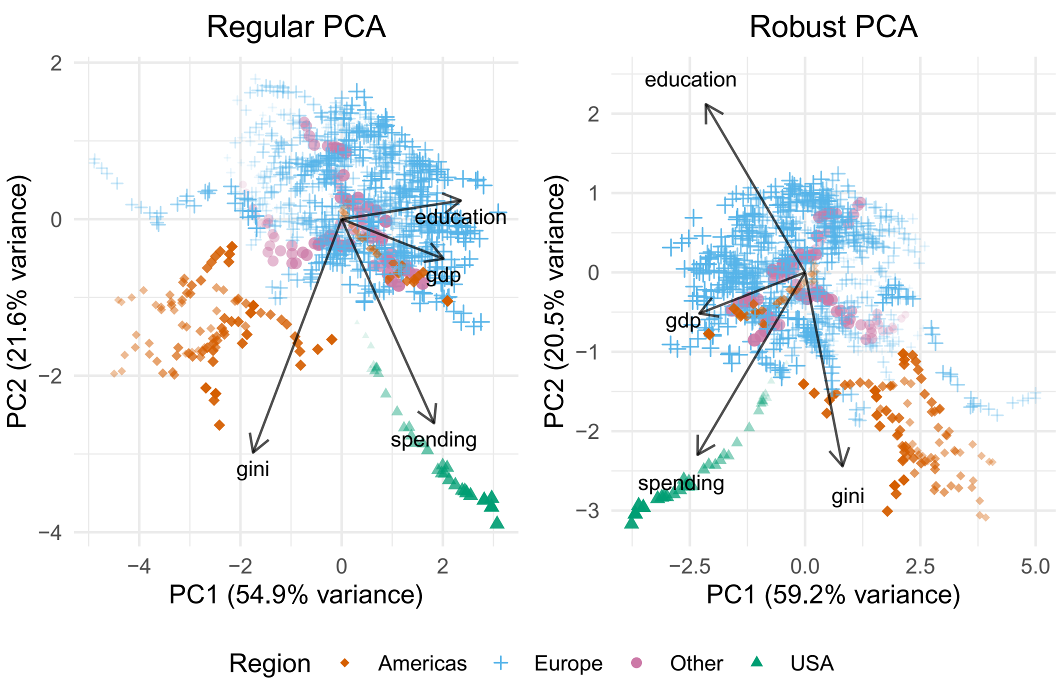

Figure 12: Biplots of the data using PCA. The loading scores are visualized by arrows. Higher opacity and larger size of the marker indicate more recent data points. "spending" denotes the spending fraction.

is explained. This means that the PCA plots can be used to make relatively reliable inferences about the countries. In addition, the first three principal components (these explain over 95% of the variation) were examined using an interactive 3D-plot. This inspection supported the conclusions from the 2D-plots.

Mahalanobis depth The Mahalanobis depth scores corroborate the findings from the principal component analysis. The regular and robust Mahalanobis scores are shown in Figure 14. In the figure, instead of all data points, the averages are used from the years 2015–2019. Despite this, the results are similar to those obtained by PCA. Again, the most recent years are not included due to the COVID-19 pandemic. Most European countries, Australia, Korea, Japan, Israel, and Canada have high scores, indicating their similarity. Central and South American countries, the US, and Turkey have low scores which supports seeing them separate from the rest. Two European countries, Luxembourg and Ireland, have the lowest scores, however. This is due to their extremely high GDP per capita in the period 2015–2019, explained by their tax haven status. Chile has also a higher score than Portugal which indicates that its socioeconomic development is approaching that of some European countries.

Notable is also the large difference between the regular and the robust Mahalanobis scores. Due to the presence of extreme outliers such as the US (healthcare spending fraction), tax havens (GDP per capita) and Central and South America (Gini), the MCD method produces results which are more consistent with the PCA.

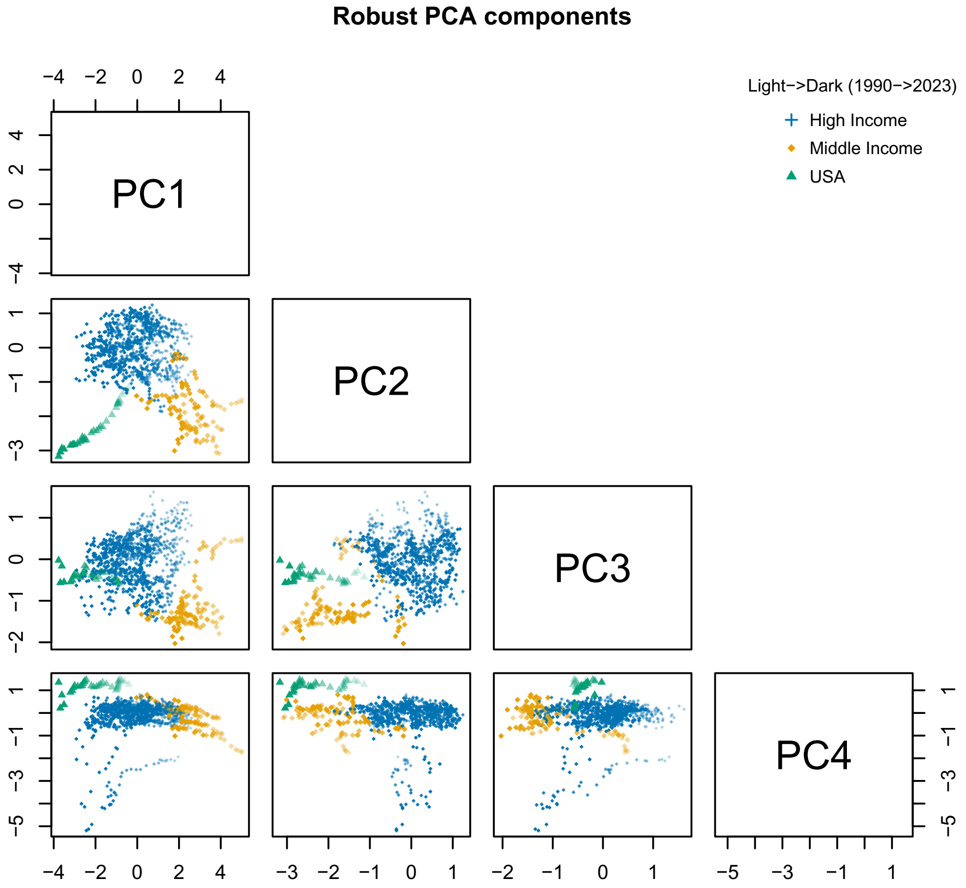

Figure 13: Pair plots of the new clusters using robust PCA.

5.3.2 Response variable

The former Soviet countries, Central and South America, and the US show the highest disease burden — Mexico and Japan are outliers

Despite the decreasing trend in age-adjusted DALY rates discussed in Section 5.1, there are large differences between countries which have largely remained the same. Figure 15 shows time series for the DALY rates across the countries. At the low end, there is a tight cluster of mostly Northern, Western, and Southern European countries as well as Korea, Australia, New Zealand, and Israel throughout the time period 1990–2023. Japan has had consistently the lowest disease burden for the whole period. The rest of the countries have had higher DALY rates and the differences are more spread out. The United States, Turkey, Central and South America, and the former Soviet countries: Lithuania, Latvia, Estonia, Hungary, Slovakia, Poland, Czechia, and Slovenia have had consistently higher rates than the tight cluster of European countries. The COVID-19 years 2020–2023 have affected countries unequally as indicated by the different-sized increases in the DALY rate. From 2015, Mexico has had the highest

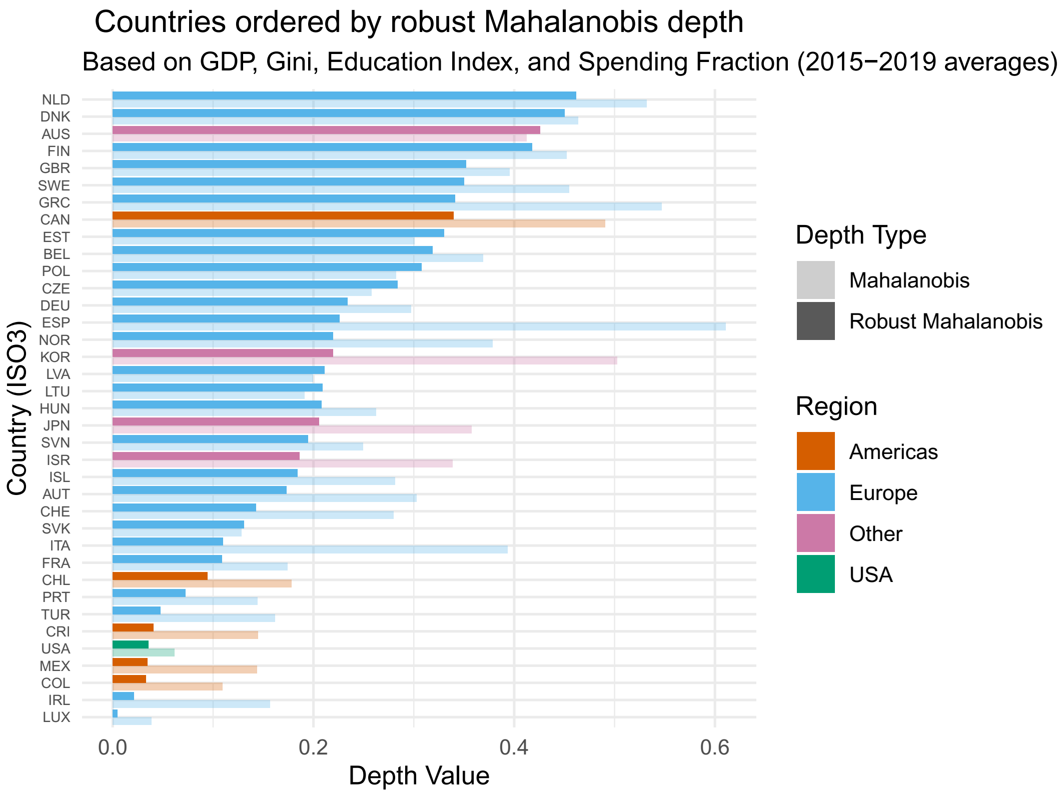

Figure 14: Regular and robust Mahalanobis scores for the countries using averages 2015–2019. New Zealand is not present due to the absence of Gini data.

DALY rates. Its increase in the DALY rate as a result of COVID-19 is also the highest among all countries.

The Mahalanobis depth scores for the 2015–2019 averages support the conclusions from the visual time series analysis. The Mahalanobis scores are presented in Figure 16 (when only the response variable is considered). Central and South America and the former soviet countries have the lowest scores (indicating the largest dissimilarity to other countries). Notable exception is, again, Chile, which upon further inspection has been able to lower its disease burden closer to that of most Western European countries. Japan and Korea also have low scores indicating their exceptionally low disease burden.

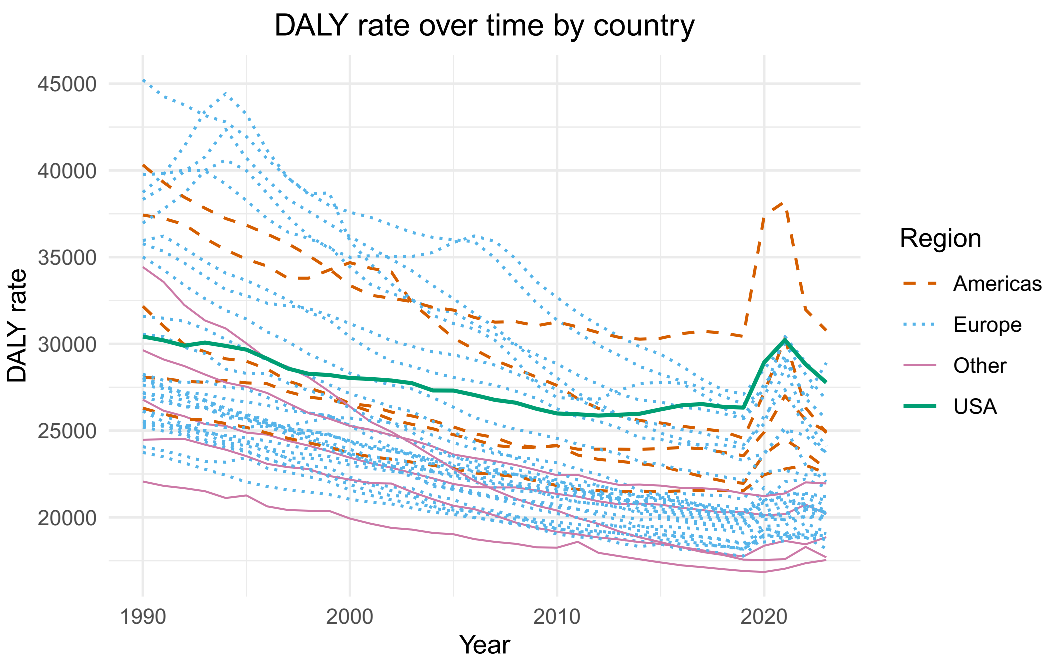

Figure 15: The age-adjusted DALY rate time series.

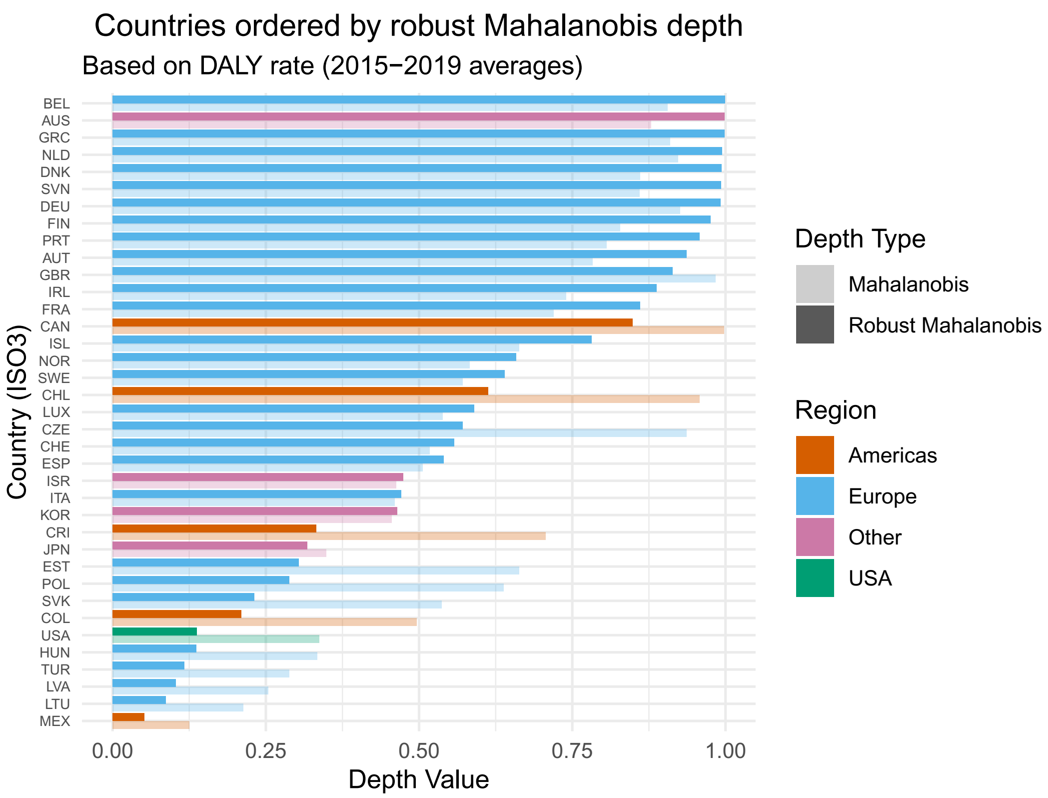

Figure 16: Mahalanobis depth scores based on the response variable.

5.4 Multiple regression

Logarithmic healthcare spending explains most of the variation in the data — the United States has a strong effect on the results

5.4.1 Univariate regression models

When comparing the univariate models, the logarithmic healthcare spending — $\log(\mathrm{SP}) = \log(\mathrm{GDP} \cdot \mathrm{SF})$ — is the best individual predictor of disease burden. It is followed closely by logarithmic GDP, which is easy to explain since spending is strongly correlated with GDP. Table 6 shows the adjusted R^2 scores of the univariate models. It also shows the fit of the models where GDP and spending are not logarithmically transformed. The lower R^2_{adj} scores imply that the returns on GDP and healthcare spending growth are diminishing (multiplicative increase is associated with a linear decrease in DALY rate).

Table 6: Univariate regression results. R^2_{adj} is the adjusted coefficient of determination of the regression model of the form $\mathrm{DALY} = \beta_0 + \beta_1 \cdot [\text{variable name}]$. Two models have been fitted for each: 1) all data and 2) the US removed.

	log(GDP)	GDP	Gini	EI	log(SP)	SP	year
R^2_{adj}	0.561	0.391	0.0616	0.219	0.582	0.367	0.196
β_1	-6780	-0.167	17124	-22559	-5093	-1.56	-256
R^2_{adj} (no USA)	0.602	0.417	0.0529	0.247	0.685	0.516	0.193
β_1 (no USA)	-7029	-0.172	16058	-23947	-5674	-2.03	-257

The logarithmic spending explains 58% of the variation in the data as a whole, which means that it is strongly associated with health. Furthermore, if the United States is removed from the analysis, the model achieves an adjusted R^2 of 69%. This highlights the fact that the United States is a strong outlier in healthcare spending. The directions of the effects of the explanatory variables on the DALY rate can be inferred from the signs of the coefficients (β_1) which are also listed in Table 6. The directions are the same as observed in the bivariate analysis. Higher GDP and healthcare spending are associated with a lower disease burden. The education index and Gini, especially, seem to be poor predictors of disease burden as indicated by R^2_{adj}. The year is almost as good a predictor as education and its negative coefficient indicates a desirable historical development. These results do not account for confounding or the combined effects of the variables (the R^2 scores sum up to over 1.0). They do, however, give a general indication of the relative importance of the different variables. All univariate models achieve high statistical significance (coefficients of the explanatory variables have p-values $< 2 \cdot 10^{-16}$).

5.4.2 The full regression model

The addition of more variables besides the logarithmic spending log(SP) does not improve the model significantly. Table 7 shows the adjusted R^2 scores of the full model (1) fitted to different sets of the data. The sets are: i) all the data, ii) all data excluding the US, and the three clusters: iii) high-income, iv) middle-income, and v) the US. Including variables besides log(SP) only increases the R^2_{adj} from 0.582 to 0.611 with all data, and from 0.685 to 0.717 with the US excluded. The clustered models fit better than the model applied to all data. In the middle-income countries, especially, the model is a very good predictor of disease burden. Despite the high R^2_{adj} score of the model fitted to the US data, the model coefficients do not achieve statistical significance.

Table 7: Comparison of the adjusted R^2 scores of the full regression model (1) applied to different sets of the data.

	all data	no USA	high-income	middle-income	USA
R^2_{adj}	0.611	0.717	0.684	0.877	0.938

The coefficients of the variables in the models, along with their respective p-values, are shown in Table 8. A large p-value indicates that there is not enough evidence to reject the null hypothesis $H_0 : \beta_j = 0$, i.e., the data does not indicate that a given explanatory variable has an effect on the response variable. Since the variables have different units and scales, the comparison of the coefficients within a model are not meaningful, except between GDP and healthcare spending. Among models fitted to different sets of the data, however, the coefficients of the same variable can be compared.

Table 8: Coefficients and p-values in the full regression model (1) fitted to different sets of the data. Coefficients which do not give reason to reject the null hypothesis $H_0 : \beta_j = 0$ at significance level 0.01 are shown in gray.

	log(SP)	log(GDP)	Gini	EI	year
All data					
β_j	-4236.8	-1380.1	2506.1	7500.8	-101.9
p-value	9.4e-12	0.06029	0.31367	4.9e-11	0.00024
No USA					
β_j	-7347.6	1455.0	-11093.9	2556.1	-26.5
p-value	< 2e-16	0.69448	5.3e-08	0.00029	8.9e-07
High-income					
β_j	-6218.2	238.4	-13071.2	8033.7	-57.6
p-value	< 2e-16	0.72	3.1e-06	1.9e-05	4.5e-06
Middle-income					
β_j	-11157.0	10171.3	16957.7	-22098.7	23.1
p-value	< 2e-16	4.1e-11	0.0055	1.1e-08	0.5670
USA					
β_j	-5095.1	1168.2	8629.6	26960.8	-2.1
p-value	0.041	0.791	0.608	0.038	0.983

The table shows interesting differences between the models. Firstly, in the model fitted to the US data, none of associations are statistically significant. This is likely due to the low number of data points. In all other models, however, log(SP) is a highly significant predictor of disease burden. Furthermore, its effect is considerably stronger when the US is excluded, further highlighting the inefficiency of the US healthcare system. The effect of healthcare spending is the strongest in the middle-income cluster (Central and South America, and Turkey).

Logarithmic GDP does not seem to be a significant predictor when the spending is included. However, in the middle-income cluster it achieves significance and has the opposite effect to healthcare spending. That is, keeping all else (most importantly, healthcare spending) constant, higher GDP is associated with higher disease burden. This is a surprising finding, but it is likely explained by Mexico which has a relatively high GDP but high disease burden. One reason for its high disease burden might be high rates of violent crime and drug abuse, which are mostly affected by factors other than healthcare spending.

In "no USA" and "high-income" groups, the effect of Gini is such that more equal countries have higher disease burden. This is explained by the Eastern European countries being more economically equal than Japan and Korea but having considerably higher disease burden. In the middle-income cluster, higher Gini is associated with a higher disease burden. Similarly to Gini, the education index seems to have opposite effects among the middle-income cluster and the rest. Surprisingly, higher education in "all data", "no USA", and "high-income" models is associated with higher disease burden. In the middle-income cluster, more education is associated with a lower disease burden.

A possible explanation for the counterintuitive findings is that the level of education and economic inequality matter when they are at low and high levels, respectively, but otherwise the differences in disease burden are better explained by other factors. Finally, the year is a statistically significant predictor in the "all data", "no USA", and "high-income" models but not in the "middle-income" model. This might reflect that there is increased knowledge about health and advancements in medicine globally, but that the middle-income countries are not able to benefit from them.

5.4.3 The strongest predictor of disease burden

Overall, the logarithmic healthcare spending seems to be the best predictor of disease burden between countries (or more accurately, year-country-pairs). It accounts for about 60% of the variation, and excluding the US, almost 70%. The explanatory variables are highly collinear, however, and the spending is almost perfectly explained by the other explanatory variables. When a model of the following form is fitted to the data:

$$\log(\mathrm{SP}) = \beta_0 + \beta_1 \cdot \log(\mathrm{GDP}) + \beta_2 \cdot \mathrm{EI} + \beta_3 \cdot \mathrm{Gini} + \beta_4 \cdot \mathrm{year} + \varepsilon,$$

it achieves an adjusted R^2 score of 0.911. Furthermore, almost all the variation in healthcare spending is explained by GDP alone. The following model:

$$\log(\mathrm{SP}) = \beta_0 + \beta_1 \cdot \log(\mathrm{GDP}) + \varepsilon,$$

with just GDP as the predictor achieves an R^2_{adj} of 0.899.

Figure 17 compares the coefficients (β — slope of the line) of log(SP) from the univariate and the full regression models applied to all data. They are the *unadjusted* and *adjusted* effects of log(SP) on the DALY rate. The figure also shows visually how good of a predictor the logarithmic spending is, and why the inclusion of the United States changes the results so drastically. The plot shows a clear decreasing trend, meaning that higher healthcare spending is associated with lower disease burden. However, the log-linearity implies that the effects are diminishing, meaning that the spending must be multiplied to yield a linear decrease.

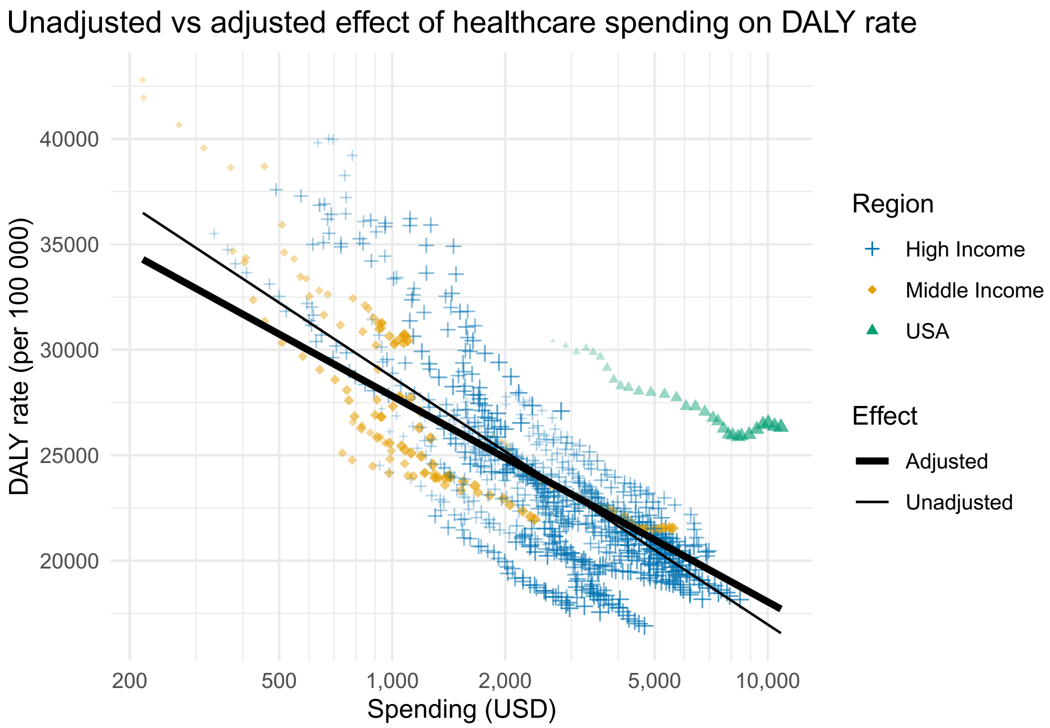

Figure 17: Unadjusted and adjusted effects of healthcare spending on disease burden. Notice the logarithmic scale of the horizontal axis. The US differs noticeably from the rest of the countries.

6 Discussion

This section explores the most important findings, discusses their implications, lists the limitations of our analysis, and gives directions for future research. First, the results reported in the last section are summarized and discussed from a more qualitative perspective. The most interesting findings are highlighted. The next subsection interprets the results further and discusses their implications. Limitations of the study are examined categorically. They are related to the methodology, the data, and the indicator variables themselves. Finally, in addition to addressing the limitations, the last subsection suggests some ways to deepen and expand the analysis.

6.1 Summary of the results

Increases in health, wealth, and education The results of our analysis show positive, i.e., desirable, development in most socioeconomic indicators in most countries. Disease burden has decreased, meaning that people are living longer and healthier. This encouraging trend is also visible in the GBD reports [2]. The real GDP per capita has more than doubled in many countries from 1990 to 2023. Similarly, the average years of education and the expected years of education for children have increased, with the largest development in the least educated countries. The proportion of GDP used for healthcare has also increased and is higher in wealthier countries which likely reflects collective willingness to spend more on health as the standard of living increases.

Unfortunately, these positive trends hide large differences between countries. There are currently almost ten-fold differences between the per-capita GDPs of countries (average of 2015–2019). Likewise, the recent disease burden of Mexico is almost twice that of Japan. If it were solely due to earlier death, the difference would correspond to a life expectancy more than a decade lower. One area without positive development is economic inequality. Gini has remained about the same everywhere throughout 1990–2023, even in high-inequality countries. Furthermore, in reality economic inequality has increased at least in Europe and the US [53]. The limitations of Gini are discussed later in this section.

Education and inequality have only minor effects A surprising finding from the regression analysis is that lower education level and higher inequality are associated with lower disease burden among high-income countries — Australia, Canada, European countries, Israel, Japan, Korea — although they only explain little of the difference in disease burden. In middle-income countries (the lowest GDPs per capita among OECD countries) the associations are the opposite — more schooling and lower inequality are linked to lower disease burden. As speculated in the previous sections, the unexpected directions of association in the high-income cluster are likely explained by Eastern Europe with its economically equal and highly educated countries that have high disease burden. An interesting observation is that former Soviet countries have higher disease burden than their nearby European neighbors. This might be caused by persistent factors related to culture or lifestyle.

These associations are not very substantial, however. The statistical significance of education and inequality as predictors does not imply that they have a large impact disease burden. In fact, they only explain an additional one to two percentage points of the variation between countries besides the logarithmic healthcare spending.

Three clusters of OECD countries There are three clusters of countries which share similar combinations of the explanatory variables. Many of these countries lie geographically close to one another, but there are exceptions. The three clusters are: 1) high-income countries including Europe, Canada, Australia, Japan, South Korea, and Israel, 2) middle-income countries including Central and South America as well as Turkey, and 3) the US. The separation between high-income and middle-income countries in wealth, economic inequality, education, and healthcare spending is to be expected, but the US radically differs from other high-income countries in regard to its healthcare spending.

When it comes to the response variable, the DALY rate, there seems to be two "clusters" of countries (see Figure 15). First, there is a tight cluster of some European countries, Japan, Korea, Canada, and Israel with low rates. The rest of the countries do not form a tight cluster, are more spread out, and all have higher disease burden. They include Southern and Central American countries, Eastern European countries, the US, and Turkey.

Healthcare spending predicts health A naive but reasonable assumption could be that the more healthy a country is, the less it needs to spend on healthcare. According to our analysis, the complete opposite holds: countries with healthier populations spend exponentially more on healthcare and health expenditure has risen as disease burden has decreased. This suggests that healthcare indeed buys health, which is the most important finding. Spending on healthcare (transformed logarithmically) explains 70% of the variation between countries (year–country-pairs) if the US is ignored (its stark difference to other countries might be explained by its healthcare system). However, equally important is the fact that healthcare spending is almost perfectly explained by GDP alone, which, in turn, is highly collinear with education and economic inequality. All variables are inherently linked, so discerning what causes what is difficult.

The United States is different A major finding of our analysis is that the United States is clearly different from other OECD countries. This is already visible in the univariate analysis, where the proportion of GDP spent on healthcare by the US is approximately double that of any other country. Accordingly, the US is clearly distinguishable from the scatter plots, and principal component analysis supports separating it into its own cluster. Conducting the regression analyses without the US increases the proportion of variation explained by the model by 10 percentage points. Most strikingly, Figure 17 shows how clearly the US "violates the law of logarithmic spending".

Investigating the reasons for the dissimilarity of the United States warrants separate research, but a possible explanation is that its fully privatized healthcare system is much less cost-effective than those of other countries. More regrettably, not only is its healthcare system less cost-effective but also less effective. In addition to the highest healthcare costs, the disease burden in the United States is much higher than in other similarly wealthy countries. Some of this might be explained by differences in lifestyle (low physical activity, unhealthy diet) but these alone do not explain such a large difference. Ironically, the public spending on healthcare in the US is also the highest in the world — according to WHO data, the US is only behind Liechtenstein in *Domestic General Government Health Expenditure (GGHE-D) per Capita in US dollars* [54].

6.2 Implications

More spending, lower returns The coefficient of log(SP) is more negative for the regression model fitted to the middle-income cluster compared to the high-income cluster (see Table 8). This suggests that the association between healthcare spending and disease burden is even "steeper" than logarithmic. That is, the gains in reduced disease burden are even smaller than logarithmic when spending is increased. Even if the association is "only logarithmic", it means that spending must be multiplied to yield a linear decrease in DALY rate. The results suggest that it is much "cheaper" for a country to go from a DALY rate of 35 000 to 30 000 per 100 000 than from 25 000 to 20 000.

The rate of change in the DALY rate with changes in spending can be calculated from the regression formula (1) (holding other variables constant) by differentiating with respect to SP as follows:

$$\frac{\mathrm{d}(\mathrm{DALY})}{\mathrm{d}(\mathrm{SP})} = \frac{\beta}{\mathrm{SP}},$$

where β is the regression coefficient associated with the logarithmic spending variable. From this:

$$\Delta(\mathrm{SP}) \approx \frac{\Delta(\mathrm{DALY}) \cdot \mathrm{SP}}{\beta}.$$

By substituting in $\Delta(\mathrm{DALY}) = -100\,000$, we can approximate how much saving a year of healthy life costs (or averting a DALY as in [12]) at a given level of healthcare spending.

For example, in a high-income country with per-capita healthcare spending at 5000 USD (and GDP per capita of 50 000 USD) the "cost of averting a DALY" is thus about 80 000 USD (using high-income cluster beta: $\beta = -6218.2$). Meanwhile, in a middle-income country with spending at 2000 USD (and GDP of 25 000 USD) the cost is only about 18 000 USD (using middle-income cluster beta: $\beta = -11\,157.0$). The result from this rough calculation is in accordance with [12] where the authors obtained a number of 69 499 USD for "very high HDI countries" and 23 782 USD for "high HDI countries".

More years of life could be saved The difference in cost-per-DALY-averted has some significant implications — especially as the trend of big health improvements with little spending is even stronger in low-income countries as suggested by our regression results (and corroborated by [12]). A grim viewpoint is that wealthy countries are essentially making a choice in saving fewer lives as they spend more on healthcare. If even a tiny fraction of what is spent on healthcare were diverted to foreign health aid, for example, comparatively much more disease could be averted. A dollar spent in the poorest countries could save up to 70 times more years of healthy live than in the richest countries according to the results from [12]. The studies evaluating specific interventions, e.g. vaccines [4] and mosquito nets [5] suggest that the difference might be even greater, and that the 80 000 USD associated with an averted DALY in the high-income countries could buy as much as 800 years of healthy life in the poorest nations.

Even when viewed through this lens, though, spending on healthcare might still be among the most justifiable investments in high-income countries — how many years of healthy life could be saved by a tiny fraction of the money spent on luxury goods or marketing arms races, for example? In very high-income countries, such as those in Western Europe and North America, the cost-effectiveness of treatments is not as carefully scrutinized as in foreign aid contexts. More emphasis is placed on what can be done rather than how to best allocate healthcare resources which are, at the end of the day, limited. While the value human life and health cannot be reduced to analyses of costs and benefits — and complete equity of health and healthcare around the world at the cost of the richest nations' own healthcare resources does not seem realistic — a more global, comprehensive, and balanced perspective can shift opinions, change policies, and consequently help eliminate more human suffering.

6.3 Limitations

6.3.1 Methodology

Our study has some methodological limitations. Most substantial are the use of age-adjusted DALY rates and the impossibility of determining causation. The latter is notoriously difficult in epidemiology, though, and much of the evidence about the dangers of smoking, for example, is correlative. Some minor limitations are the use of t-tests in significance testing, which assume that the residuals are normally distributed, which might not perfectly hold in our data, and possible endogeneity in the model — among other things, disease burden can affect the productivity of the workforce, and thus GDP. In addition, we excluded the COVID-19 years 2020–2023 from the regression analysis. To better understand the differences between the effects of the pandemic in countries and the reasons for those differences, methods different from ours should be used. Disease burden caused by COVID-19 has already been extensively studied, e.g. in [33].

DALYs in an aging population The age-standardization of DALYs makes comparisons between countries valid as the effect of different age distributions is removed. A country with a large proportion of elderly people will have higher disease burden than a country with a younger population, even if all individuals live equally healthy and long lives in both countries (because old people, on average, have higher disease burden compared to younger adults). Age-standardization controls for this by comparing the DALY rates in respective age groups and then assuming that both countries have the same age distribution. For comparing healthcare costs, however, this is not ideal. In an aging population — where the proportion of the elderly increases — it should be expected that the per-capita healthcare costs increase although the age-adjusted DALY rate stays the same. Thus, a country with an older population will have to spend more on healthcare compared to a country with a younger population, even if the same resources for each age group are offered in both countries. This skews our results; countries with a younger population "get an unfair advantage". Further research could be conducted using crude DALY rates instead of the age-adjusted rates.

Causation For some interventions such as vaccines, their effect on reducing disease burden is provably causative as they are required to be tested in controlled experiments for effectiveness against the disease. Thus, it can be reliably estimated how much a vaccination program reduces disease and that the reduction is specifically due to the vaccines. Furthermore, when the costs are known, the cost of an averted DALY can be accurately estimated. Our analysis does not come even close to establishing such a link between healthcare spending and health outcomes. When interpreting our results and especially their interpretations in Section 6.2 it is important to note that correlation does not necessarily equal causation. Different methodologies are required to show causation between increases in general healthcare spending and decreases in disease burden.

6.3.2 Indicators

Many of the indicator variables we chose for our analysis offer an imperfect measurement of the phenomenon they are trying to quantify. For example, Gini has not fully captured the increasing economic inequality in Western Europe and North America [53]. The reasons for this are beyond the scope of our work, but it is possible, for example, that people lie on the household surveys that the calculation of Gini is based on, or that a sufficient number of ultra-wealthy individuals is not reached, meaning that the surveys do not represent the whole population faithfully. Still, we chose to use Gini because of its popularity and ease of understanding. Similarly, GDP is a flawed measure of a nation's wealth. Outflowing profits are not subtracted from GDP, despite that economic value not going towards the well-being of the country's citizens. Instead, gross national income could have been used such as in HDI, for example. Finally, a minor limitation of education index is the way that it is calculated; if mean schooling and expected schooling increase beyond 15 and 18 years, respectively, the index does not increase. This might explain the observed nonlinear relationship between GDP and EI. The former can increase limitlessly but the latter cannot.

6.3.3 Data

There is missing data on the socioeconomic indicators (see Table 4). Especially many data points are missing in the Gini data and the health spending data. It is possible that if the missing data existed, it would change the results. Even if no data were missing, there would still not be enough to make statistically significant inferences about individual countries (as can be seen with a model fitted to the US only in Table 8). We use linear interpolation for filling in missing data, but more advanced data imputation methods could be tried in future research. Although the data is from reliable sources (e.g. the World Bank and the United Nations), there is uncertainty in some values and completely accurate reporting is impossible. Lastly, there are uncertainty intervals reported with the DALY data which are not taken into account in our analysis.

6.4 Further research

Analysis of more countries The easiest way to expand the analysis is to conduct it globally. OECD countries constitute less than a fifth of the number of countries in the world and a much smaller proportion of the global population. The Global Burden of Disease study provides data on almost all countries, and GDP and education data is similarly widely available. The results would show whether the same associations hold in considerably less wealthy countries besides the OECD.

Explaining remaining variation While healthcare spending explains about 70% of the variation in disease burden (if the US is excluded), there is 30% of unexplained variation, and it is not due to differences in education or economic inequality. A possible reason is the difference in healthcare systems, which is a likely explanation for the dissimilarity of the US. Differences in lifestyle (diet and physical activity), environment (pollution), and culture are other potential sources of the remaining variation. Historical context dating further back than 1990 might also affect disease burden. At any rate, it would be useful to understand what other factors are associated with health as addressing them could improve population health more cost-effectively compared with increased healthcare spending.

Breakdown of DALYs and healthcare costs Finally, to make the analysis more detailed, DALYs could be divided into injuries, NCDs, and CMNNs or even further into cardiovascular disease, cancer, mental illness, etc. It is possible that the differences in disease burden between countries are mostly due to differences in some specific causes, e.g. mental illness. Likewise, breaking down the healthcare spending is required to understand what drives the increasing costs in OECD countries and why the effect of spending on disease burden is diminishing. The analysis could shed light on the areas where increased resources would bring about the most benefits.

7 Conclusions

Higher spending on healthcare is clearly associated with lower disease burden. It alone explains most of the variation in disease burden among OECD countries. Healthcare spending itself, however, is almost perfectly explained by GDP. Education or economic inequality do not strongly predict disease burden when differences in healthcare spending are accounted for. The effect of healthcare spending seems to be diminishing, meaning that as disease burden decreases the costs increase exponentially. This finding is supported by the literature, indicating that averting the loss of years of healthy life is the more expensive the richer the country is. While our analysis does not prove causality, it shows a strong link between healthcare spending and disease burden, and the causative effects of many specific health interventions are supported by the literature. Potential future research topics include expanding the analysis to more countries, finding factors which explain the remaining variance, incorporating the breakdown of disease burden (injuries, communicable, and noncommunicable diseases) into the analysis, as well as scrutinizing the healthcare costs.

References

[1] R. Bonita, R. Beaglehole, and T. Kjellström. *Basic epidemiology*. World Health Organization, 2006.

[2] GBD 2023 Disease and Injury and Risk Factor Collaborators. "Burden of 375 diseases and injuries, risk-attributable burden of 88 risk factors, and healthy life expectancy in 204 countries and territories, including 660 subnational locations, 1990-2023: a systematic analysis for the Global Burden of Disease Study 2023". en. In: *Lancet* 406.10513 (Oct. 2025), pp. 1873–1922.

[3] A. M. Utami et al. "Economic evaluation of COVID-19 vaccination: A systematic review". en. In: *J. Glob. Health* 13.06001 (Jan. 2023), p. 06001.

[4] S. Ozawa et al. "Cost-effectiveness and economic benefits of vaccines in low- and middle-income countries: a systematic review". en. In: *Vaccine* 31.1 (Dec. 2012), pp. 96–108.

[5] J. Wisniewski et al. "Systematic review and meta-analysis of the cost and cost-effectiveness of distributing insecticide-treated nets for the prevention of malaria". en. In: *Acta Trop.* 202.105229 (Feb. 2020), p. 105229.

[6] S. Sarkar et al. "Cost-effectiveness of HIV prevention interventions in sub-Saharan Africa: A systematic review". en. In: *EClinicalMedicine* 10 (Apr. 2019), pp. 10–31.

[7] E. Marseille et al. "Scaling up integrated prevention campaigns for global health: costs and cost-effectiveness in 70 countries". en. In: *BMJ Open* 4.6 (June 2014), e003987.

[8] D. Kostova et al. "The cost-effectiveness of hypertension management in low-income and middle-income countries: a review". en. In: *BMJ Glob. Health* 5.9 (Sept. 2020), e002213.

[9] K. Singh et al. "Cost-effectiveness of interventions to control cardiovascular diseases and diabetes mellitus in South Asia: a systematic review". en. In: *BMJ Open* 8.4 (Apr. 2018), e017809.

[10] S. F. Hope et al. "A systematic review of economic evaluations of population-based sodium reduction interventions". en. In: *PLoS One* 12.3 (Mar. 2017), e0173600.

[11] A. Tuvdendorj, T. Feenstra, and E. Buskens. "Cost-effectiveness of four tobacco control interventions in Mongolia". en. In: *Nicotine Tob. Res.* 25.11 (Sept. 2023), pp. 1719–1726.

[12] R. Daroudi et al. "Cost per DALY averted in low, middle- and high-income countries: evidence from the global burden of disease study to estimate the cost-effectiveness thresholds". en. In: *Cost Eff. Resour. Alloc.* 19.1 (Feb. 2021), p. 7.

[13] OECD. *The OECD: Better policies for better lives*. https://www.oecd.org/en/about.html. Accessed: 2025-11-24.

[14] World Bank. *Poverty and Inequality Platform: Methodology*. https://data.worldbank.org/indicator/SI.POV.GINI. Accessed: 2025-11-10.

[15] M. O. Lorenz. "Methods of measuring the concentration of wealth". In: *Publications of the American statistical association* 9.70 (1905), pp. 209–219.

[16] World Bank. *Poverty and Inequality Platform: Methodology*. https://datanalytics.worldbank.org/PIP-Methodology/. Accessed: 2025-11-10.

[17] P. Bondarenko. *gross domestic product*. In: *Encyclopaedia Britannica*. URL: https://www.britannica.com/money/gross-domestic-product (visited on Dec. 17, 2025).

[18] United States Bureau of Foreign and Domestic Commerce. *National Income, 1929–1932*. Prepared with contributions from the Seventy-Third Congress (1933–1935) and Simon Kuznets. Washington: U.S. Government Printing Office, 1934.

[19] United Nations, Department of Economic and Social Affairs, Statistics Division. *GDP, Per Capita GDP at constant 2015 prices - US dollars*. https://unstats.un.org/unsd/snaama/Basic. Accessed: 2025-12-17. 2025.

[20] OECD. *Incorporating estimates of household production of non-market services into international comparisons of material well-being*. Tech. rep. Oct. 2011.

[21] P. Bondarenko. *gross national income*. In: *Encyclopaedia Britannica*. URL: https://www.britannica.com/money/gross-national-income (visited on Dec. 17, 2025).

[22] K. Montevirgen. *GDP price deflator: Nominal output vs. real output*. In: *Encyclopaedia Britannica*. URL: https://www.britannica.com/money/gdp-price-deflator (visited on Feb. 11, 2026).

[23] S. Eldridge. *purchasing power parity*. In: *Encyclopaedia Britannica*. URL: https://www.britannica.com/money/purchasing-power-parity (visited on Feb. 11, 2026).

[24] UNDP. *Human Development Index (HDI)*. https://hdr.undp.org/data-center/human-development-index. Accessed: 2025-12-16.

[25] D. T. Jamison et al. *Disease control priorities, third edition (volume 9)*. en. Ed. by D. T. Jamison and H. Gelband. 3rd ed. Washington, D.C., DC: World Bank Publications, Dec. 2017.

[26] C. J. L. Murray and A. D. Lopez. "The global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020". en. In: (1996). Published by the Harvard School of public health on behalf of the World Health organization and the World Bank. Distributed by Harvard University Press.

[27] World Health Organization. *WHO methods and data sources for global burden of disease estimates 2000–2019: Global Health Estimates Technical Paper WHO/DDI/DNA/GHE/2020.3*. Technical Paper WHO/DDI/DNA/GHE/2020.3. Accessed 2025-11-24. Geneva, Switzerland: World Health Organization, Department of Data, Analytics, Division of Data, Analytics, and Delivery for Impact, Dec. 2020. URL: https://www.who.int/docs/default-source/gho-documents/global-health-estimates/ghe2019_daly-methods.pdf.

[28] GBD 2023 Causes of Death Collaborators. "Global burden of 292 causes of death in 204 countries and territories and 660 subnational locations, 1990-2023: a systematic analysis for the Global Burden of Disease Study 2023". en. In: *Lancet* 406.10513 (Oct. 2025), pp. 1811–1872.

[29] Brecht Devleesschauwer. *Introduction to DALYs: historical and technical basis*. European Burden of Disease Network. Sept. 10, 2021. URL: https://www.burden-eu.net/docs/ts2/01-01-bod-concepts.pdf (visited on Feb. 6, 2026).

[30] J. A. Salomon et al. "Disability weights for the Global Burden of Disease 2013 study". en. In: *Lancet Glob. Health* 3.11 (Nov. 2015), e712–23.

[31] GBD 2019 Diseases and Injuries Collaborators. "Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019". en. In: *Lancet* 396.10258 (Oct. 2020), pp. 1204–1222.

[32] G. 2. F. Collaborators. "Burden of disease scenarios for 204 countries and territories, 2022–2050: a forecasting analysis for the Global Burden of Disease Study 2021". In: *The Lancet* 403.10440 (2024), pp. 2204–2256.

[33] C. Chen et al. "Global, regional, and national characteristics of the main causes of increased disease burden due to the covid-19 pandemic: time-series modelling analysis of global burden of disease study 2021". en. In: *BMJ* 390 (July 2025), e083868.

[34] GBD 2021 Nervous System Disorders Collaborators. "Global, regional, and national burden of disorders affecting the nervous system, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021". en. In: *Lancet Neurol.* 23.4 (Apr. 2024), pp. 344–381.

[35] C. J. L. Murray. "The Global Burden of Disease Study at 30 years". en. In: *Nat. Med.* 28.10 (Oct. 2022), pp. 2019–2026.

[36] GBD 2023 Demographics Collaborators. "Global age-sex-specific all-cause mortality and life expectancy estimates for 204 countries and territories and 660 subnational locations, 1950-2023: a demographic analysis for the Global Burden of Disease Study 2023". en. In: *Lancet* 406.10513 (Oct. 2025), pp. 1731–1810.

[37] Institute for Health Metrics and Evaluation (IHME). *Global Burden of Disease 2023: Findings from the GBD 2023 Study*. Seattle, WA, 2025.

[38] M. K. Seo, P. Baker, and K. N.-L. Ngo. "Cost-effectiveness analysis of vaccinating children in Malawi with RTS,S vaccines in comparison with long-lasting insecticide-treated nets". en. In: *Malar. J.* 13.1 (Feb. 2014), p. 66.

[39] P. J. Neumann et al. "A systematic review of cost-effectiveness studies reporting cost-per-DALY averted". en. In: *PLoS One* 11.12 (Dec. 2016), e0168512.

[40] M. S. Syeed et al. "Pneumococcal vaccination in children: A systematic review and meta-analysis of cost-effectiveness studies". en. In: *Value Health* 26.4 (Apr. 2023), pp. 598–611.

[41] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2021. URL: https://www.R-project.org/.

[42] Institute for Health Metrics and Evaluation. *GBD Results*. https://vizhub.healthdata.org/gbd-results/. Accessed: 2025-11-10.

[43] International Monetary Fund. *GDP per capita, current prices*. https://www.imf.org/external/datamapper/PPPPC@WEO/OEMDC/ADVEC/WEOWORLD. Accessed: 2025-11-10.

[44] United Nations. *Human Development Reports*. https://hdr.undp.org/data-center/documentation-and-downloads. Accessed: 2025-11-10.

[45] OECD. *Health spending*. https://www.oecd.org/en/data/indicators/health-spending.html. Accessed: 2025-11-10.

[46] K. V. Mardia. *Multivariate Analysis*. en. Ed. by K. V. Mardia. Probability and mathematical statistics. San Diego, CA: Academic Press, Dec. 1979.

[47] R. Y. Liu. "Data depth and multivariate rank tests". In: *L1-statistical analysis and related methods* (1992), pp. 279–294.

[48] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. en. Nashville, TN: John Wiley & Sons, Dec. 1987.

[49] F. E. Harrell Jr. *Regression modeling strategies*. en. 2nd ed. Springer Series in Statistics. Basel, Switzerland: Springer International Publishing, Aug. 2015.

[50] B. H. Baltagi. *Econometric Analysis of Panel Data*. en. Cham, Switzerland: Springer Nature, Mar. 2021.

[51] J. G. MacKinnon and H. White. "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties". In: *Journal of econometrics* 29.3 (1985), pp. 305–325.

[52] M. Ezekiel. *Methods of Correlation and Regression Analysis*. New York: John Wiley & Sons, 1930.

[53] T. Piketty and E. Saez. "Inequality in the long run". en. In: *Science* 344.6186 (May 2014), pp. 838–843.

[54] WHO. *Global Health Expenditure Database*. https://apps.who.int/nha/database/Select/Indicators/en. Accessed: 2026-01-21.

A Appendix: AI use statement

Generative artificial intelligence systems have been used in making this thesis. More specifically, the free versions of OpenAI's ChatGPT and Anthropic's Claude have been used between September 2025 and March 2026. The specific models are not always disclosed to the user but can include the top models GPT-5.2 (ChatGPT) and Sonnet 4.5 (Claude). The applications have been: 1) topic ideation (finding gaps in the literature), 2) code generation and editing, especially for formatting tables and graphs, 3) LaTeX-formatting, and 4) finding literature sources (articles, books, etc.). For the last application, the tools have been used only for finding sources, not for summarizing them or generating text. All cited sources have been read by the author.