Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Geography

Marte Sortland

The effect of accessible and high quality drinking water on stunting

A first step of considering drinking water into the physical capacity curve

Master's thesis in Globalisation and Sustainable Development
Supervisor: Thomas Halvorsen
November 2022

NTNU
Norwegian University of Science and Technology

Marte Sortland

The effect of accessible and high quality drinking water on stunting

A first step of considering drinking water into the physical capacity curve

Master's thesis in Globalisation and Sustainable Development
Supervisor: Thomas Halvorsen
November 2022

Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Geography

NTNU
Norwegian University of
Science and Technology

Abstract

The thesis analyses the effect of access and quality of drinking water on stunting. The results are then used to discuss the inclusion of drinking water into the theory of the physical capacity curve. For the analysis, linear regression is used to test the effect of four variables for access and quality of drinking water on stunting. Interaction terms are used to test for interdependency between the variables for drinking water and minimum acceptable diet. The hypotheses that are made focus mainly on the results of accessible drinking water and high quality drinking water.

The results show that accessible drinking water decreases the level of stunting. The interaction terms reveal that there is interdependence between minimum acceptable diet and each of the variables piped drinking water, limited drinking water, and unimproved drinking water. Based on limited previous literature and the results from this thesis it is not sufficient to conclude on the inclusion of drinking water into the physical capacity curve, although the results strengthen the argument in favour of an inclusion. This thesis highlights the need of further analyzing the effects of drinking water on stunting, focusing on the interdependence of drinking water and nutrition. This thesis also emphasizes the need of further research to consider including drinking water as a part of the physical capacity curve.

Preface

Writing this thesis has been a challenging but inspiring period, gaining incredible insight into a specific field of development economics.

I am thankful for the support and feedback from my supervisor Thomas Halvorsen. Being available to answer any questions have helped me steadily move forward, in clarifying both big and small issues. Most importantly, his support on my topic has echoed and kept me motivated even in the slower periods.

I would also like to express my deepest gratitude to Professor Colin Green. His lectures in development economics were key for my enthusiasm about this topic. Professor Colin Green was also a great help in finding literature on the specific theme of my thesis.

A special thank you goes to my friends Nils Arne Lindaas, Frida D. Vang, Fredrik Nesse, and Thea Myklebust. I am grateful for the advice and constructive criticism, as well as all the encouragement they have provided me in the commentary and elsewhere. Even the smallest remarks you have given me has been greatly appreciated.

Lastly, I am incredibly grateful to everyone who has cheered me on and encouraged me every step of the way. My thesis has been the one and only topic of conversation for a while now, but thank you for listening. Moral support should not be underestimated and has kept me going especially in times of frustration. When the going gets tough, the tough gets going, but that is not without a big network of people who in some ways have expressed their support.

Content

Tables

Figures

List of abbreviations

JME: the Joint Malnutrition Estimates

JMP: the WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene

SDG: Sustainable Development Goals

UNICEF: the United Nations Childrend's Fund

WASH: Water, Sanitation and Hygiene

WB: the World Bank

WHO: World Health Organization

1. Introduction

The theory of the physical capacity curve is used to describe the effect of income, operationalized as nutrition, on the work capacity of humans (Dasgupta, 1997; Ray, 1998). The physical capacity curve fails to include drinking water, which is essential to sustain human life and maintain a minimum quality of life. Nutrition is therefore arguably not the only factor with a substantial effect on work capacity. Accessible and safely managed drinking water is of vital to keep hydrated, which in turn is vital to sustain cells and the ability for bodily processes such as sweating (Popkin, D'Anci, & Rosenberg, 2010, p. 441). Thus, it is reasonable to believe that the access to safely managed drinking water is significant for the work capacity of humans.

The topic of drinking water is relevant because more than 25% of the world's population lack the access to safely managed drinking water (WHO, 2021, pp. 7-8). The lack of access to safely managed drinking water is problematic because it poses a risk for diseases and infections, the deterioration of bodily functions and consequently productivity (Edokpayi et al., 2018; Haq, Mustafa, & Ahmad, 2007). The topic of drinking water is a global issue and has been raised in the United Nations, specifically through dedicating a Sustainable Development Goal (SDG) to water in SDG 6: access to safe drinking water and adequate sanitation (UNEP, 2021, p. 87).

1.1. Background

The physical capacity curve omits drinking water as a factor that affects work capacity. Although there is no nutritional value or energy gained from drinking water, the body requires water to effectively absorb the nutrients and energy from food (Choudhary, Schuster, Brewis, & Wutich, 2021). Drinking water is a basic need, and the interconnectedness of nutrition and water can be illustrated as a machinery where nutrition is the fuel and drinking water is the oil ensuring that the machine works smoothly.

There is a limited number of studies linking access and quality of drinking water specifically to work productivity. In the assessment of the importance of safely managed drinking water, Jain (2012, p. 1) has stated that access to safely managed drinking water is linked to the alleviation of poverty and enhancing productivity. Conversely, the lack of access to good quality drinking water has been linked to lower worker productivity (Haq et al., 2007). These studies link drinking water to work productivity and poverty based on the effect of drinking water on public health. While safely managed drinking water is free from contamination, alternative drinking water sources have a higher probability of being contaminated. Both microbiological and chemical contamination have severe negative impacts on health (Edokpayi et al., 2018; Jain, 2012). With microbiological contamination such as bacteria, it is more frequent with diarrhoea, which in the most severe extent may cause death by dehydration (Todaro & Smith, 2012). Chemical contaminants also gives incapacitating illness, which varies from acute nausea and vomiting to fetal abnormalities (Hunter, MacDonald, & Carter, 2010). These health effects cause a lower worker productivity, which in some circumstances may lead to poverty.

Several studies point to a more frequent water-borne diseases as a consequence of not having access to high quality drinking water. This is highlighted due to the severe implications of diseases, such as malnourishment and dehydration (Choudhary et al., 2021; Todaro & Smith, 2012). Young children are especially vulnerable, and in developing countries and rural areas without access to high quality drinking water, diarrhoea is a leading cause of child mortality (Edokpayi et al., 2018, p. 15; Hunter et al., 2010). On the other hand, having access to safely managed drinking water reduces the mortality caused by water-borne diseases by 70% (Haq et al., 2007, p. 1137). It is crucial that people have access to safely managed drinking water to ensure the coverage of their basic needs and to prevent diseases. Diseases caused by a contaminated drinking water source has the potential to become widespread in a short amount of time, posing a risk to societies as a whole (Jain, 2012).

The objective of this thesis is to test the effect of access and quality of drinking water on stunting. Stunting is a negative health outcome of having insufficient access to nutrition or nutritional uptake, which causes a stunted growth in children (Black et al., 2013; WHO, 2014). Choudhary, Schuster, Brewis and Wutich (2021) have studied the effect of drinking water on stunting in India. Their study found that inaccessible drinking water is directly linked to higher levels of stunting, as well as indicating an indirect effect through diet diversity. Their findings are supported by others, pointing to stunted growth and fetal abnormalities as consequences of ingesting polluted drinking water (Edokpayi et al., 2018; Lang, 2015). In order to test the effect of drinking water on stunting this thesis uses multiple linear regression with data for over 200 countries.

The effects of drinking water on health outcomes are acknowledged, but there is a knowledge gap in the effect of drinking water on productivity. This thesis will contribute to promote the importance of drinking water, especially in the context of global development. Drinking water and nutrition are both basic needs that are required to sustain human life (WHO & UNICEF, 2021). Additionally, the consumption of contaminated drinking water can diminish the nutritional value of food, through increasing the probability of diseases (Choudhary et al., 2021). It is therefore problematic that nutrition alone is mentioned in the theory of the physical capacity curve. Although the main objective is to establish if access and quality of drinking water has an effect on stunting, this thesis also aims at discussing whether the findings support the argument of including drinking water into the physical capacity curve.

1.2. Research questions

The aim of the study is to assess if there is a relationship between drinking water and stunting. The results will then contribute to discuss if the effects are strong enough to support the inclusion of drinking water into the theory of the physical capacity curve. In order to do so this thesis will answer three research questions:

1. *Is there an association between accessibility of drinking water and stunting?*
2. *Is there an association between quality of drinking water and stunting?*
3. *Is the relationship between drinking water and stunting strong enough to support inclusion of drinking water into the physical capacity curve?*

1.3. Structure

Chapter two presents the theoretical framework for this paper. The chapter starts by giving a thorough explanation and a definition of the key concepts of drinking water and stunting. The chapter then focuses on the existing theories regarding the interdependence of food and water. Next, the theory of energy balance is presented. This establishes the foundation for the physical capacity curve. Lastly, the physical capacity curve and its implications are explained in depth.

Chapter three introduces the methodology. The chapter starts by giving a general approach to the study, the research questions and presenting the hypotheses. Following this, is an introduction of the data. This includes a description of the datasets, the reliability and validity, critique to the data as well as adjustments that have been made prior to the analysis. An explanation of the variables and their measures follows this section, which introduces the dependent variable, four independent variables for different types of drinking water sources, and independent controlled variables. Lastly, an assessment of the preconditions for the analysis is made in addition to describing the analysis approach.

Chapter four contains the results of the analysis. The level of significance is established in the beginning of the chapter. The results are then presented categorically by the research questions.

First, the results of the effect of accessibility of drinking water on stunting are presented. Next, the results focus on the effect of drinking water quality on stunting. Lastly, the results of the interaction terms of available nutrition and drinking water are presented.

Chapter five consist of the discussion. The chapter begins with a brief commentary on the models from the analysis. The results are then discussed using the theoretical framework presented in chapter two. First, the effect of accessibility on drinking water will be discussed. Next is the effect of high quality drinking water on stunting. Following this is a discussion of whether drinking water should be included in the physical capacity curve based on the findings. The chapter then provides a summary of the discussion and an overview of the outcomes of the hypotheses. Lastly, an assessment of the weaknesses is provided in the critique of the study.

Chapter six provides a conclusion of the significant findings in the study for each of the research questions. After the summary, final comments are presented. Finally, suggestions are made for further research on this topic.

2. Theory

This chapter will expand on the theoretical framework that is the motivation for this paper. The chapter will start by giving thorough definitions of the concepts of *drinking water* and *stunting*. Next is an account for the existing literature on the interdependence of nutrition and water. This part will highlight the research of the development economic perspective. Following this is an account for the theory of energy balance which provides a basis for the physical capacity curve. Lastly, the general objective is to modify the physical capacity curve, and this theory will therefore be thoroughly explained.

2.1. Defining drinking water

It is a fact that water is the most basic requirement to sustain life and health, and consequently recognized as a human right (Edokpayi et al., 2018; Haq et al., 2007; WHO & UNICEF, 2021, p. 4). There is a global consensus on the importance of securing the demand for drinking water for all. The issue is actualized through the SDG 6, with the specific Target 6.1 to provide universal access to safe drinking water (UNEP, 2021, p. 87; WHO & UNICEF, 2021, p. 1). The Target 6.1 is important to consider due to it giving an account of what is considered essentially good drinking water. Although most of the planet is covered by water, not all is safe or available to drink (Hunter et al., 2010; Jain, 2012). This proves the need of a measurement of what is considered good drinking water for humans.

Drinking water refers to water that is collected with the intention of consuming it as drinking water. Similarly, drinking water sources are the points where drinking water is collected (WHO & UNICEF, 2021). Drinking water sources excludes the places of origin, such as dams or water reservoirs, and refers solely to the point where the consumer collects the drinking water. The target of universal access to safe drinking water assumes that the drinking water source is on the premises of the household with water free from contamination (Choudhary et al., 2021; WHO & UNICEF, 2021). A benefit of the SDG Target 6.1. is that it provides a measure of optimal drinking water conditions and subsequently how neglecting these conditions can cause people to experience suboptimal drinking water.

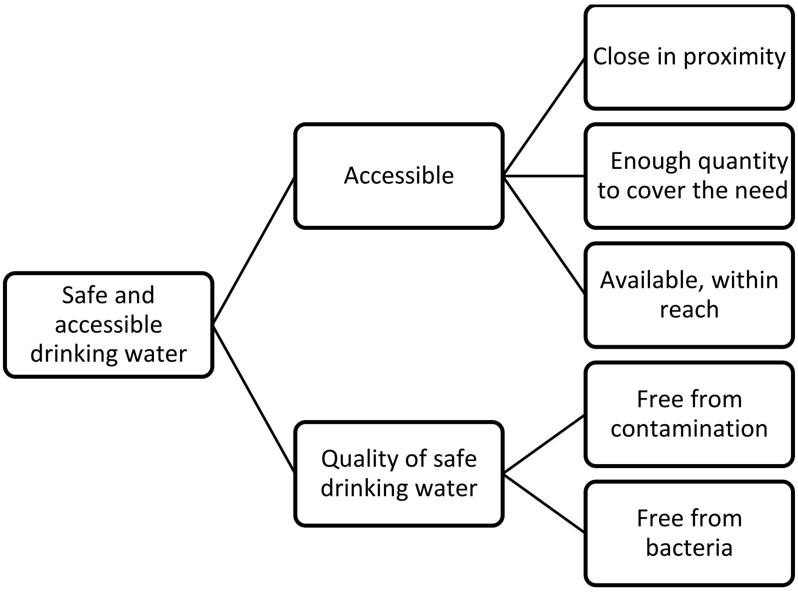


Figure 2.1.: The elements of safe and accessible drinking water.

The SDG Target 6.1. explains access to safe drinking water as a twofold issue, as illustrated in figure 2.1. It consists of the issue of making drinking water accessible as well as providing safe water, indicating a certain level of quality of the drinking water. Access to drinking water consists of three major conditions. First, it has to be in close proximity to consumers, which according to the Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP) refers to sources located on the premises of the household (WHO & UNICEF, 2021). This implies that the consumers live right next to their drinking water source or are connected to a water transmission network, such as piped water into their household. In these situations, the drinking water is immediately accessible for the individual consumer. Next, there has to be an adequate drinking water quantity to cover the needs of the consumers (Choudhary et al., 2021; Jain, 2012). Having sufficient drinking water quantity means that the household has sufficient drinking water to supply their daily use. Lastly, the drinking water source must be available in the sense that it is possible to collect drinking water from a reachable and stabile source. To achieve this the source must 1) not be physically hindered, and 2) give individuals access to water supply schemes (Haq et al., 2007, p. 1145). Additionally, the source must be stable and not destroyed or frequently damaged. This relates especially to areas prone to extreme weather that physically damage infrastructure (Luh et al., 2015). An unreliable source of drinking water cannot be considered to be accessible if the supply is not guaranteed and there exists periods of inaccessibility.

What is stated as quality in figure 2.1. refers to whether the water is safe to drink or not (Choudhary et al., 2021; de França Doria, 2010). The quality and safety of the drinking water is determined by the existence of external contamination, focusing especially on faecal matter and chemical contamination (Jain, 2012; WHO & UNICEF, 2021). Water with a high level of contamination is not considered safe to drink due to the possible negative health outcomes that may follow from ingesting water with either bacterial or chemical contaminants (Todaro & Smith, 2012). The quality can also be affected by public perception, which Jain (2012) identifies as physical contamination, such as a change in appearance or sensory experience like smell. While sensory perception can be useful for individuals to determine the quality of their drinking water, it requires knowledge of what is high quality as a reference. Nevertheless, the aesthetic of drinking water heavily affects the perception of quality even when there is no actual health risk (de França Doria, 2010). Due to the uncertainty of actual risk, aesthetics and sensory experiences will not be a determinant for the actual quality of drinking water.

The body's water intake is mostly covered by drinking water or other beverages containing water. Studies from the United States suggest that only around 20% of the daily water intake comes from food (Popkin et al., 2010; Sawka, Cheuvront, & Carter, 2005, p. 32). This percentage may differ based on diets with varying levels of water-rich foods, but the majority of the water intake is still based on drinking water and beverages (Popkin et al., 2010). Thirst and hunger are not equal bodily needs, so it is reasonable to believe that one cannot live off water from food alone, thus highlighting the importance of access to safe drinking water. However, there is no consensus on what the daily required water intake for the individual is. Factors such as physical activity, physical and cognitive energy expenditures, and thermoregulation through for example sweat are considered to have a significant impact on the required total water intake (Popkin et al., 2010). Thus, higher levels of physical activity, general energy expenditures and thermoregulation causes increased dehydration and a need to replace more water than what lower levels of physical activity, general energy expenditures and thermoregulation require.

2.2. Stunting

The term stunting refers to the impaired growth and development in children due to malnutrition, infections in the child's early years, and also being deprived of psychosocial stimulation (WHO, 2014, pp. 1-2) A child is considered stunted if their height-for-age is more than two standard deviations below the WHO Child Growth Standards median (WHO, 2014, p. 1). This will also be the foundation for the operational definition of stunting. The height of children and their stunted counterparts generally follows the same trend, as shown in figure

2.2., differentiating by gender. The height for girls and boys is based on measures in environments for optimal growth, giving the assumption that the stunted children live in conditions that are not optimal for normal development.

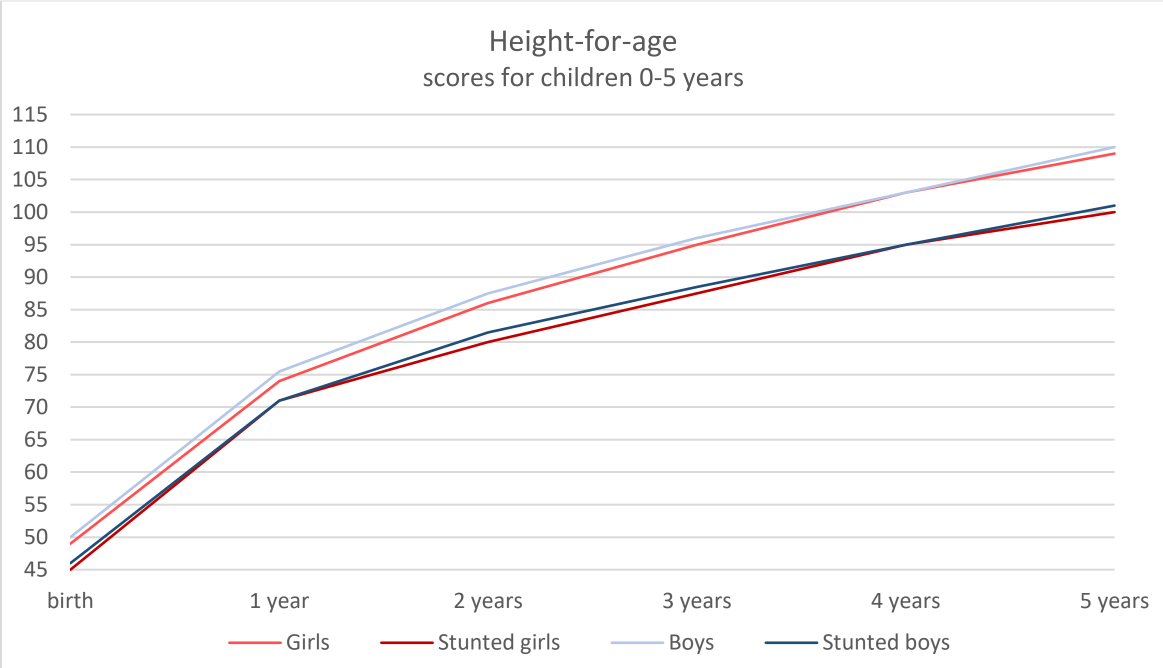


Figure 2.2.: Height-for-age scores for children between 0-5 years old (WHO).

The main contributing factor to stunting in children is malnutrition, in the form of a limited dietary intake, variation or no nutritional intake at all during pregnancy as well as in the early developmental years (Black et al., 2013; UNICEF, WHO, & WB, 2019; WHO, 2014). A preliminary assumption is that for the development of a child, it is essential that the mother has adequate storage or continuous intake of nutrients to support the development of the fetus during pregnancy. Black et. al. (2013) support this assumption, adding that the nutritional status from as early as the time of conception is significant for a fetus' growth and development. It provides a reasonable point, as the nutritional needs of the mother and fetus must both be covered from the time of conception. In the case of adolescent pregnancy, maternal growth is still ongoing, causing competition for the nutritional intake that becomes available (Black et al., 2013; WHO, 2014). Limited storage or intake of nutrition, indicated for example as a low BMI, can therefore cause one or both parties to experience a shortage that hinders optimal development. Moreover, parental stunting increases the likelihood of stunting in their offspring (Black et al., 2013, pp. 428-429). Thus, maternal stunting can be a contributing factor to childhood stunting.

During a child's early years, it is crucial to cover their nutritional needs for their development and long-term health (Black et al., 2013, p. 427). By contrast, malnutrition and undernutrition is widely recognized for causing stunting and underdevelopment (UNICEF, a; Vilcins, Sly, & Jagals, 2018; WHO, 2014). Another factor frequently mentioned is the significance of repeated infection in the first 1000 days of the child's life (WHO, 2014, p. 1). Repeated infections refer to infections causing a decrease in the ability to absorb nutrients or in any other way leading to undernutrition (Vilcins et al., 2018, p. 556). The first 1000 days of a child's life are crucial for their development, and although these infections can cause negative health outcomes for shorter time periods, they have lasting effects because they happen in this time period (Black et al., 2013). The Global Nutrition Report (2021) also includes psychosocial stimulation as a determinant for stunting, but this is not explained further in reports or literature. Assuming it refers to cognitive development, this is already associated with impaired physical growth and development (WHO, 2014). Ultimately, there exists little evidence suggesting that psychosocial stimulation impacts stunting on an equal level to crucial factors such as nutrition and repeated infection.

Taking the determinants for stunting into consideration, a suitable operational definition is the impaired growth of a child, where the measured height-for-age is more than two standard deviations below the WHO Child Growth Standards median, caused by chronic malnutrition and infections affecting nutritional intake. Because of the emphasis on nutrition as a determinant for stunting, it is a suitable measure, or proxy, for child malnutrition. On a national level, the proportion of children suffering from stunting gives a clear suggestion of the nutritional status because it provides an indication of chronic undernutrition (UNICEF, 2019). Stunting is prevalent in situations with nutritional deficits in times where covering the nutritional demand is critical, giving reason to believe that the general nutritional status is a status of limited access to nutrition.

2.3. The interdependence of food and water

There are few established theories within the social sciences regarding the relationship between drinking water and food in humans (Engell, 1988, p. 134). Similarly, there are few empirical studies regarding this relationship. The studies that do exist have contradicting conclusions and seemingly few participants, making them problematic to generalize for entire populations (McKay, Belous, & Temple, 2018). Intake of both food and drinking water is known from natural science to be vital for humans, thus we can propose the theory that food and drinking water have to be mutually present to sustain life.

All foods contain nutrients, although it varies in recommended intake, and the concept of food is used for energy intake through nutritional intake (McKay et al., 2018, p. 17). Subsequently, the amount and types of nutrients makes up the total nutritional status of an individual. Regular drinking water does not contain any nutrients, making it problematic to include into the measure of the nutritional intake of an individual. Nutrients provide energy to the body while drinking water fail to do so due to the lack of nutritional value, supporting the exclusion of water in theories tied to energy and nutritional intake. However, Popkin et al. (2010, p. 446) state that water quantitatively is the most important nutrient. With this statement, two ideas are proposed. First, water is essential and the amount of drinking water in an individual's diet should exceed the amount of food. As discussed in chapter 2.1. regarding drinking water, there is no definite requirement for water intake. However, this point provides an impression of the ratio between water and food intake. Secondly, water is regarded as a nutrient, or at least of equal importance as nutrition. Although it does not hold any nutritional value, water should still be considered on the same grounds as nutritional intake. Supporting this is the argument that the body requires water in order to efficiently absorb the nutrients found in food (Choudhary et al., 2021, p. 171). This presents water and nutritional intake as equally valuable, and more importantly mutually dependent to sustain the human body. By this explanation, water would be linked to all theories regarding nutrition and the body, such as energy balance in the body, because it is a facilitator for nutritional intake.

2.4. The theory of energy balance

The theory of energy balance is based on the first principle of thermodynamics, stating that the amount of energy that exists is constant. The energy cannot be destroyed, only moved between organisms through the gain, loss or storage of the energy (Hill, Wyatt, & Peters, 2013, p. 111). The theory of energy balance refers to a state within the body where energy intake equals the energy expenditure (Dasgupta & Ray, 1990, p. 196; Hill et al., 2013). Energy balance is the balancing of the energy that is gained, lost and stored within the body. A characteristic of energy balance is stability in body weight. An energy imbalance would conversely cause weight gain or weight loss, depending on if there is a larger proportion of energy gained or lost, respectively (Hill et al., 2013).

For the sake of consistency, I will use the defining features outlined by Debraj Ray (1998) to further explain the concept of energy balance. These features are based on previous literature on the topic by Dasgupta and Ray (1986, 1990), Ray and Streufert (1993), and Ray (1992), and lays the foundation for the physical capacity curve. Ray argues that within the theory of energy balance there are four main components. These are: energy input, resting metabolism, energy output, and the body's ability to store and borrow energy (Ray, 1998, pp. 272-274).

The *energy input* encompasses all the energy that is gained by the body. The energy is gained mainly through the consumption of food and calorie intake (Dasgupta & Ray, 1990; Ray, 1998; Ray & Streufert, 1993). All consumption that provides the body with energy is considered an energy input. Taking this a step further, Ray also argues that access to food in most cases is equivalent to income, particularly in the poor population (Ray, 1998). Because food is an item that requires a certain capital or asset to acquire, this assumption is reasonable to bear in mind.

Resting metabolism is the energy that is spent on basic bodily functions and can be described as actions that naturally exist to sustain the body. This concept is also known as resting metabolic rate (RMR) (Hill et al., 2013, p. 111). Resting metabolism consists of involuntary actions such as maintaining body temperature, sustaining heartbeat and respiratory systems, and providing the minimum requirements of resting tissue (Ray, 1992, 1998). It is also argued that it includes minimum voluntary actions such as eating and minimum sanitation (Ray & Streufert, 1993, p. 65). There is a major difference in how much energy individuals spend on voluntary actions. It is therefore reasonable to assume that voluntary actions only cover a minimum requirement to sustain the body. The majority of the energy input goes toward these actions of resting metabolism (Dasgupta, 1997; Ray, 1998; Ray & Streufert, 1993). Controlling the energy required to cover resting metabolism proves difficult due to the bodily functions mostly being involuntary actions. What is certain is that resting metabolism is a constant expense.

The energy output is the energy that is required for performing actions such as labor or physical activities. Energy output includes all energy that is required for activities beyond resting metabolism, although the emphasis is on work (Ray, 1992, 1998; Ray & Streufert, 1993). This means that although energy output can be energy lost by walking up a set of stairs, the intention is to describe the energy used to perform physical labor. It can be problematic making an estimate for what is generally required to cover the energy output because it varies greatly. Ray and Streufert (1993, pp. 65-66) estimate for 900 calories (Kcal) per day for an active male farm worker. While Ray (1998, p. 273) cites an estimate of 400 Kcal per day for moderate work, he does criticize this measure and points out that hard physical labor such as tree felling can account for as much as 502 Kcal per hour. Type of work is important because it highlights that physical labor is typically done by the poor (Ray, 1992, 1998). Ultimately, energy output increases with hard physical labor because it requires more energy.

The body's ability to store and borrow energy is the process of adjusting to energy excesses or deficits caused by an imbalance between energy input and the sum of resting metabolism and energy output (Ray, 1992, 1998). Borrowing refers to the body's ability to use stores from the body during energy deficits (Ray, 1998). This provides an explanation to why one could go long before eating, such as during fasting periods. Borrowing energy in times of energy deficits causes a decline in body mass (Ray, 1992). Because there is a limit to the body's stores there is a limit to what an individual can borrow. Therefore, long-term energy deficits have vital consequences like the breakdown of the body by causing incapacitating debility and death (Ray, 1992; 1998, pp. 273-274). Conversely, the body's ability to store energy is the ability to save any energy excesses, causing an increase in body mass (Ray, 1992, 1998). Although some of the energy excess may dissipate, a continuous energy imbalance of excess energy input may provide an explanation for the development of overweight and obesity.

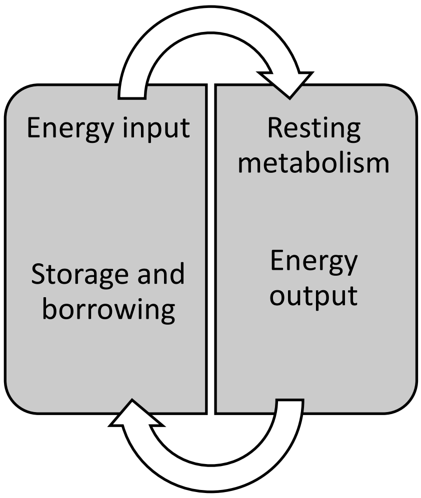


Figure 2.3: the theory of energy balance as a continuous cycle with energy entering and exiting the body.

Energy balance is a continuous process, and figure 2.3. visualizes how the four components of energy balance play into this process. Energy input or stores of energy provide the energy to cover resting metabolism and energy output. What remains, if any, is then stored and can be used in times of energy deficits.

With energy being gained, lost and moved within the body, the theory of energy balance can also be explained as an equation. For the equation Dasgupta and Ray (1990; 1992) include energy metabolism, which is the process of converting nutritional intake into energy.

(1) $$\delta_t x_t = r_t + q_t - b_t, \quad t \geq 0$$

In this explanation, δ is the efficiency of energy metabolism at time *t*. This measure is between 0 and 1, with a low value representing a lower efficiency of energy metabolism (Dasgupta & Ray, 1990, p. 200). x is the energy input for the individual. Next, r is the resting metabolism or resting metabolic rate. q is the energy output, the energy spent on physical activities beyond the resting metabolism. lastly, b is the energy that is stored or borrowed within the body. The energy stored or borrowed can be positive or negative, with borrowing causing the measure to be positive, and storing energy causing it to be negative (Ray, 1992). For it to be considered an energy balance, this equitation must be true, meaning that the left side must equal the right side. There may exist more complicated relationships, but this explanation includes the essential aspects of the theory of energy balance.

2.5. The physical capacity curve

The physical capacity curve is a term used to describe the modelled relationship between income and work capacity and assumes that all income is converted to nutrition (Ray, 1998, pp. 274-275). This relationship is a theory developed by Dasgupta and Ray based on the efficiency wage hypothesis by Harvey Leibenstein (Dasgupta & Ray, 1986; Jha, Gaiha, & Sharma, 2009; Leibenstein, 1978). The efficiency wage hypothesis argues that at low levels of nutrition, usually in developing countries, workers are not physically able to do hard manual labor. This causes low productivity, low wages, low purchasing power and consequently poor nutrition (Jha et al., 2009, p. 982). Dasgupta and Ray (1986; Ray, 1998) use this hypothesis to systematize the relationship between income and work capacity into a graph, visualized in figure 2.4.

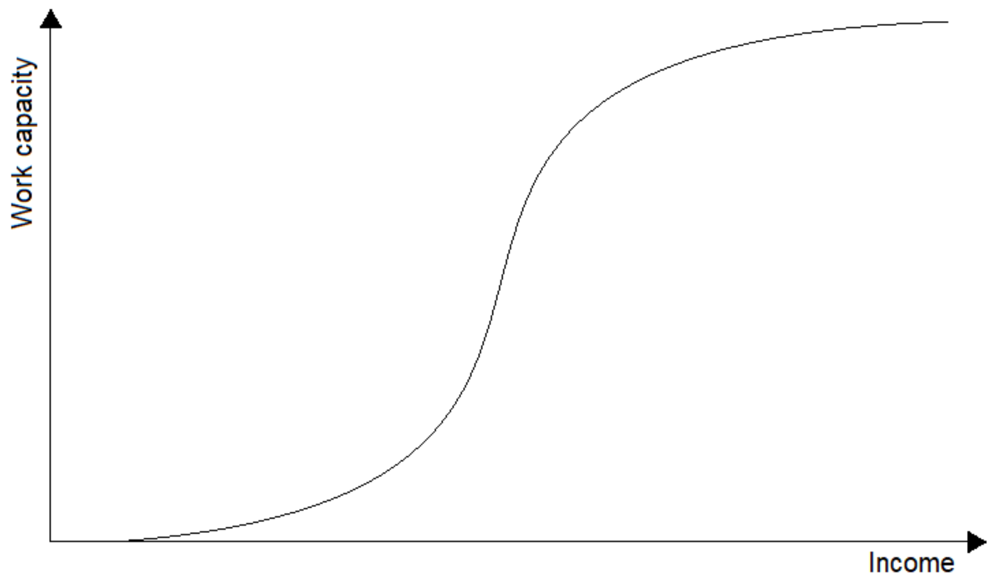


Figure 2.4.: The physical capacity curve, modelled after Ray (1998, p. 274).

The basic assumption for the physical capacity curve is that all income is used on consumption in the form of nutrition, that is food (Dasgupta, 1997; Ray, 1998). Ray (1998, p. 489) argues that even though a more realistic scenario would be one where 70% of income is spent on nutrition, the assumption does not cause less substance to the theory.

(2) $$Income = Nutrition$$

The physical capacity curve in figure 2.4. shows the relationship between work capacity and income as nutrition. As income increases, so does work capacity, making it a positive relationship. Thus, an individual has more capacity to do work-related tasks when their income and nutritional intake increases. Using the theory of energy balance as a foundation, the curve visualizes the increase in possible energy output that is demanded to execute tasks if energy input increases.

One can identify three sections within the physical capacity curve, which are visualized in figure 2.5. below. Initially, low levels of income have no significant effect on work capacity. Within this stage, all of the income goes toward covering resting metabolism and there is little or no additional income to exceed the needs of the resting metabolism (Ray, 1998, p. 275). Thus, the income is not sufficient to cover any additional energy outputs, and therefore has no effect on the work capacity for the individual. The physical capacity curve is thus shifting upwards, even at low income. At what is defined in figure 2.5. as middle income, a marginal increase in income leads to a more significant change in work capacity. As income increases, a surplus above satisfying the resting metabolism is created. The energy which is required to cover resting metabolism is fulfilled, and surplus income can go into work capacity. At the higher levels of income, the resting metabolism is covered and there is sufficient energy input to also cover energy output. At this level, it is not necessary to spend the entire income on food, consequently giving any additional income limited effects on work capacity as shown in figure 2.5., where the line flattens. Similar to diminishing returns, there cannot be an infinite increase in work capacity because of the natural bodily limits that exist.

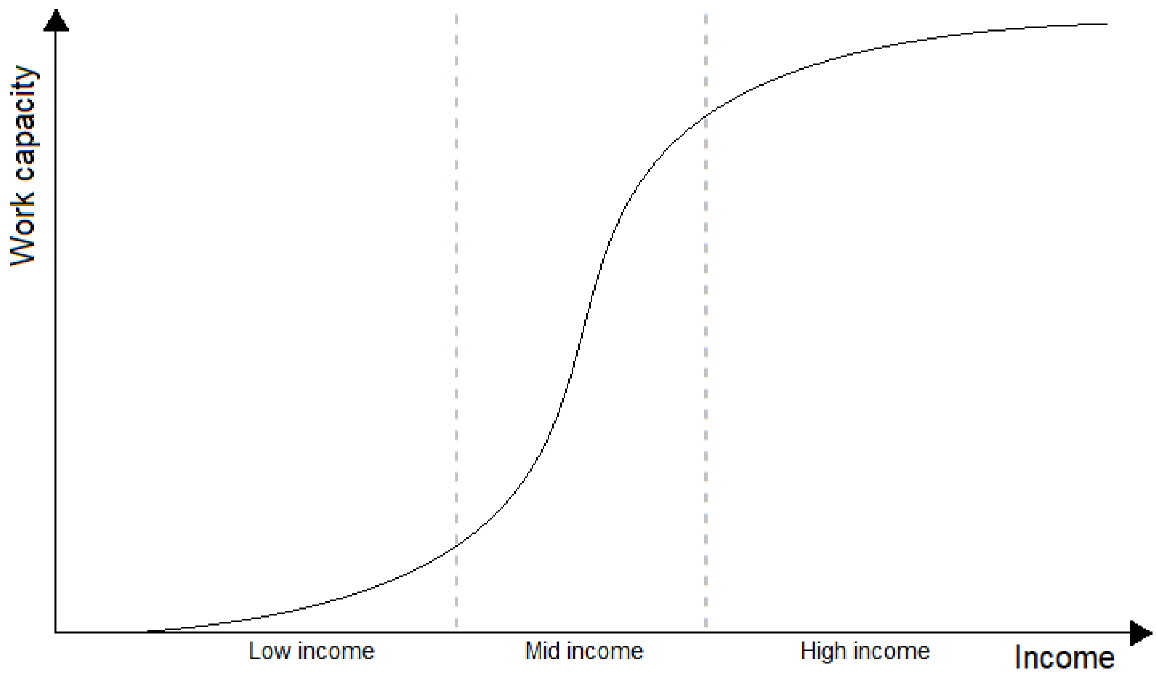


Figure 2.5: The three sections in the physical capacity curve, interpreted.

There is an underlying circular relationship in the physical capacity curve. The level of income implies how much nutrition the individual potentially can obtain which affects the possible energy output and the individual's work capacity. Although work capacity is an ambiguous measure, it indicates at what rate the body can convert the energy within it into physical work (Dasgupta, 1997). Work capacity then determines the ability of an individual to execute tasks that generate a monetary income as illustrated in figure 2.6. below. The level of income, in turn, indicates the possibility of obtaining a healthy nutritional status which again impacts the work capacity as shown in figure 2.6 (Ray, 1998). While the physical capacity curve is the curve describing the relationship between income and work capacity, this underlying circular causality will in this paper be called the physical capacity circularity. The physical capacity circularity differs from the initial efficiency wage hypothesis by Leibenstein because it does not assume a direction of the relationships, only that there exists a circular causality.

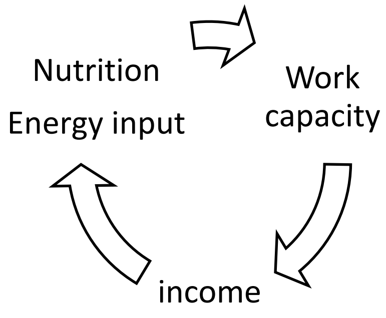


Figure 2.6.: The cycle implied in the physical capacity curve.

Taking the physical capacity circularity in figure 2.6. into account, one can identify possibilities similar to the three sections in the physical capacity curve in figure 2.5. At low levels of nutritional intake, work capacity remains low and does not create the opportunity of increased income (Ray, 1998, pp. 274-275). This relationship can be deemed an evil circle as illustrated in figure2.7. and relates to poverty trap described in the efficiency wage hypothesis (Jha et al., 2009).

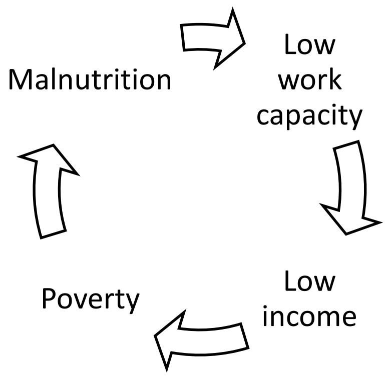


Figure 2.7.: Possible self-reinforcing cycle for a low-income low-work capacity relationship.

As a contrast, at high levels of income it is assumed that the energy input exceeds the resting metabolism and energy output combined. Consequently, the individual has a surplus of energy that increases their possible work capacity (Ray, 1998). Participating in the labor market at all and having the work capacity to increase their capability to do work may generate a higher level of monetary income. The increased income can be used to increase energy input which has a positive effect on work capacity, as is prevalent in the section of mid income in *figure 2.5*.

When a high level of income is achieved the physical capacity circularity works to maintain a sufficient nutritional status and high work capacity.

The physical capacity curve does not consider the consequences of overweight or obesity. This is a potential effect of increased income, where the individual can afford a higher energy input than energy output, causing an increased body mass (Ray, 1998). In addition to natural bodily limits, a body mass above the healthy weigh-for height can contribute to diminishing returns as income increases. However, the aim of the authors is to focus on the economies of developing countries and the possibility of poverty traps, and I will therefore not consider this scenario further.

3. Methodology

This chapter will expand on the methodology used in this paper. The chapter starts by briefly explaining the general research approach. Next is an account of the approach to the research questions followed by the research questions with their respective sets of hypotheses. Following is a presentation of the datasets that are used. This includes a description of the datasets and collection methods, comments on reliability and validity for the data used in this paper, critique of the data, and what adjustments have been made prior to the analysis. After this, the variables will be introduced and explained. This follows a natural order of first presenting the dependent variable *stunting*, followed by the four independent variables of different types of drinking water sources, and lastly is a presentation of the controlled variables. Next is an assessment of the preconditions for the analysis, focusing on providing solutions for possible problematic results. Lastly, the analysis approach will be described and the software that is used for this paper will be stated.

3.1. General approach

This paper uses a combination of inductive and deductive research approaches. The motivation for this paper is to suggest a modified version of the physical capacity curve as well as proving the relationship between drinking water and nutrition. Based on the already existing theories, three research questions with hypotheses have been formulated. These are presented below, in chapter 3.3. The hypotheses will be tested using empirical data to either confirm or reject the hypotheses, and finally draw a general conclusion.

3.2. Approach to research questions

There are two main objectives of this paper. The concrete objective of this study is to establish what effect the access to safe drinking water has on childhood stunting as a proxy for nutrition. A more general objective is to determine if the effect of drinking water on stunting is strong enough to support an inclusion of drinking water in the physical capacity curve, on similar grounds as nutrition.

In chapter two, the close relationship between stunting and nutrition was established. Stunting can be used as a proxy for nutrition because it is a result of chronic undernutrition or malnutrition. Conversely, it is reasonable to assume that stunting is not as frequent in areas with sufficient nutrition available. In reference to the literature on the interdependence of food and water, it also implicitly means that stunting is not as frequent in areas with sufficient drinking water access and quality. For the general objective to include drinking water into the physical capacity curve, the analysis will prove if drinking water is a determinant for stunting and thus should be considered as nutrition on the same grounds as food. Stunting is limited to a lower height-for-age in children up to five years old, but it reflects the general living conditions with a special regard to nutrition.

The databases have data for countries for several years, providing panel data. One key advantage of using panel data is that it contains more information about variability/variation, and thus better captures any potential effects. To make better use of the time aspect, it is reasonable to use a dummy approach for years to include how time contributes to explaining the variation of the dependent variables.

3.3. Research questions and hypotheses

Chapter one presented the three research questions this paper intends to answer. In order to answer the research questions, sets of hypotheses have been formulated. The hypotheses reflect the expected outcomes, based on the theoretical framework presented in chapter two.

Research question 1: Is there an association between accessibility of drinking water and stunting?

H1: Accessible drinking water causes a decline in stunting.

H2: Piped water significantly reduces the probability of stunting.

Research question 2: Is there an association between quality of drinking water and stunting?

H3: High quality of drinking water causes a decline in stunting.

H4: Safely managed drinking water significantly reduces the probability of stunting.

H5: Safely managed drinking water has the most significant effect on stunting.

Research question 3: Is the relationship between drinking water and stunting strong enough to support inclusion of drinking water into the physical capacity curve?

H6: Accessible drinking water and available nutrition has a greater effect on stunting than accessible drinking water alone.

H7: High quality drinking water and available nutrition has a greater effect on stunting than high quality drinking water alone.

H8: There is a pattern between the accessibility and quality of drinking water and stunting.

3.4. Data

3.4.1. About the datasets

This paper uses several datasets from UNICEF Data Warehouse and the World Bank (WB) Databank. Both provide free and publicly accessible data, with data on more than 500 indicators covering key topics such as health, demography, economy and education.

The Child Malnutrition dataset is provided by the Joint Malnutrition Estimates (JME), a working group consisting of UNICEF, WHO and WB. The dataset consists of updated estimates for stunting in children on a national-, regional- and global level. The aim of the JME is to provide a measure for the SDG indicator for child nutrition under Target 2.2 (UNICEF, WHO, & WB, 2021, p. 1). The data is collected through household surveys as primary data which in turn is used to create national estimates. Each of the collaborating agencies may also supplement this process by using their own approaches to obtain data. This involves using existing networks, technical support or data source catalogues (UNICEF, WHO, & WB, 2021). The results are posted annually given there is sufficient data, with some years having multiple observations. For the years with multiple observations, the most recent observation has been chosen for that year. A total of 204 countries are included, 155 of which provide national data source such as household surveys or administrative data. The remaining 49 countries have a modelled estimate that is only included in regional or global aggregates (UNICEF, WHO, & WB, 2021).

Water, sanitation and hygiene (WASH) is a collaborative global monitoring project between WHO and UNICEF, through the JMP. The aim is to provide a solid database for the SDG 6, specifically Target 6.1 and Target 6.2 (WHO & UNICEF, 2018, p. 4). As indicated by the name, the dataset monitors and measures the progress in water, sanitation and hygiene at the household level, in schools and healthcare facilities. These are in turn published as measures on a national level. The data is collected by JMP from various sources such as national statistical offices, administrative data, household surveys, or from already existing and available datasets from research institutes, regional initiatives or data catalogues (WHO & UNICEF, 2021). The results are posted annually. WASH includes all 232 countries or territories included in the 2015 version of the UN Population Division's World Population Prospects. However, the three areas Caribbean Netherlands, Holy See, and Western Sahara do not have available data for any of the WASH-indicators (WHO & UNICEF, 2018, p. 8). For countries with missing values, estimates are only made for regional or global aggregates and not on a national level. I will not be using the WASH aggregates and therefore not take the JMP aggregation method into concern.

The World Development Indicators is a database provided by WB. The database contains over 1000 indicators, which includes indicators that specifically examine the SDGs. Data is available for 217 economies as well as regional aggregates (WB, a). There are multiple collection methods for each indicator, such as data collection from national statistical offices, household surveys, and already available datasets from established institutions The results are posted annually, and adjusted for fiscal years, making the published result reflect the actual year of collection (WB, 2018).

3.4.2. Reliability and validity

The datasets are administered and published by well-known/established global actors. There is no reason to suspect that the publishers have a personal or business-driven agenda influencing the quality of the data. Additionally, the methodology and metadata used is publicly accessible, securing transparency. I am confident in their methods and data published.

The Child Malnutrition dataset measures stunting in children, which differs from the other variables measuring the total population. This is not considered an issue in this case, as the other variables provide a contextual information regarding the conditions that the children live in. The age-group from the Child Malnutrition dataset is also included as it captures a crucial period of development, in which this effect can be recorded more clearly.

I have chosen data for the years 2014-2019, and data is available for all the years in this time period. This period was chosen because it is the most recent data not affected by the COVID-19 pandemic. Thus, the chosen data measures a situation of "normalcy" and reflects the reality in a more general sense, increasing the validity of the results. The pandemic would be an extraordinary event that could provide extreme values not representative for the general situation in a country. Moreover, it is a step to secure reliability of the data because the COVID-

19 pandemic may have prevented the possibility to conduct in-person household surveys or measurements due to social restrictions. Using data from 2014-2019 secures that the relationship between stunting and water access and quality is analyzed, rather than this relationship during a global pandemic.

3.4.3. Critique of data

The datasets are published annually. While they do provide regular data, they do not show regular seasonal differences. This is important to note, especially for the WASH dataset because access to water and quality to water may vary drastically across seasons. Previous literature has mentioned the impacts of seasonality in drinking water access, for example the consequences of seasonal weather on drinking water sources (see Jain, 2012, p. 2 and Luh et al., 2015). In years with multiple observations the observations happen sporadically rather than regularly for the seasons.

Another challenge with the WASH dataset is that there are 12 indicators of drinking water sources. Although this can contribute to providing more nuanced information, it becomes more difficult to identify an indicator for best fit in a simplified model of *have* and *have-nots* for accessibility and quality. The indicators are not mutually exclusive, so one risks overlaps in any combinations not used by UNICEF.

Although household surveys are a part of the data collection, the data is only available on a national level. There are not found any differences in distribution of drinking water on a micro level. This makes the findings prone to level errors because it assumes that data on national level is generalizable to the individual level.

Lastly, during the process of data collection errors can occur. There will always be a risk of error during collection of data and the inputting of data, as well as the possibility of wrongfully exclude respondents. However, because there are multiple actors running extensive reviews of the data, I do not consider it to pose a big risk to the results.

3.4.4. Adjustments to the data

Some preliminary adjustments have been made to the data prior to the analyses.

The data has been imported to the text-editing software Notepad++ before doing the analyses to remove excessive elements that do not contribute to the analyses, such as comments, sources and references, and duplications of indicator name. In the event of multiple observations during the same year, only the most recent measure was kept while older ones were removed.

The dependent variable and the independent variables used have been retrieved from a different database than the controlled variables. This caused a problem in merging the datasets because names of countries as well as country codes differed. This was the case with for example the country code for Latvia, where UNICEF used country code "LAT" and WB used country code "LVA", or in instances where names differed slightly such as Czechia and the Czech Republic. This was altered to prevent double observations of the same countries and did not affect the data further.

Although the topic of gender is extremely interesting, it is further from my research question and aim for this paper. Therefore, in the cases of three or more observations for one year (one for female, one for male, one for total), I have removed the gender-specific observations, keeping a measure for the total population. While I do recognize the impact gender may have on water and nutrition, it is better suited to include in further research.

3.5. Variables

3.5.1. Dependent variable

Stunting is the dependent variable in this paper and is used as a proxy variable for nutritional well-being. The variable is a continuous variable, measured as the percentage of children between 0-5 years old who fall below -2 standard deviations from the height-for-age median of the reference population (UNICEF, a). Because of the previously discussed connection between nutrition and stunting, where it is well established that malnutrition causes stunting, *stunting* serves its purpose well as a proxy variable for the nutritional well-being in the population.

The variable measures below -2 standard deviations from the median. Thus, it includes both moderate (-2 standard deviations) and severe forms (-3 standard deviations) of stunting. The definition of the variable holds the same meaning as the operational definition previously stated in chapter 2.2. and is defined by UNICEF as a condition resulting from chronic or recurrent undernutrition both during the pregnancy in-utero and in early childhood (UNICEF, 2019; UNICEF, WHO, & WB, 2019).

3.5.2. Independent variables

To cover all the potential nuances of accessible and safe drinking water, I have chosen four independent variables. Each variable describes a scenario of accessibility and quality outcomes. To simplify, model 3.1 below describes the potential outcomes of accessibility and quality in terms of have and have-nots. Even though it is a simplification of the situations, it contributes to providing a more nuanced image of how the accessibility and quality of water affects stunting. Although they are chosen to best present the different situations of accessibility and quality, the four independent variables are not mutually exclusive. For example, piped water can be both systems of well maintained infrastructure as well as faulty pipes with leaks.

The distinction of quality is made based on the variables that best represents the essence of their category. High quality drinking water derives from high quality drinking water sources. Conversely, undetermined quality drinking water stems from unregulated sources that have a higher probability of carrying contaminated drinking water. Consequently, there is no substantial difference between the drinking water source and the drinking water itself.

All of the four main independent variables are continuous and measured as a percentage (a score between 0-100). They have been kept at this measurement level when imported as a dataset.

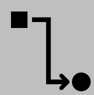
	High quality	**Undetermined quality**
Accessible	Safely managed drinking water 	Piped drinking water sources 
Not accessible	Limited drinking water services 	Unimproved drinking water source 

Table 3.1: Four variables to cover the differences in quality and access to drinking water.

Safely managed drinking water shows the percentage of the population who access water from an improved source located on premises, available when needed and free from faecal and chemical contamination (UNICEF, c; WHO & UNICEF, 2021). Improved drinking water sources refers to sources with advanced water treatment technologies to prevent outside contamination. Due to this, there may be variations in the quality, but the systems in place make it more likely that the water is safe to ingest than water from unimproved sources (WB, 2022). Like other infrastructure there may occur faults or accidents, but generally the *safely managed* *drinking water* is a measure of the proportion of the population who have drinking water free from contamination immediately accessible, for example as treated water from tap on their premises.

Limited drinking water is a measure of the percentage of the population that have access to an improved drinking water source, but where collection time takes 30 minutes or more for a round trip including any potential queuing (UNICEF, b). For this variable, improved sources of drinking water can either occur naturally or as a result of active protection, such as a public standpipe or protected dwell (WB, 2022).

Piped drinking water is the percentage of the population who use piped drinking water sources (UNICEF, d). Piped drinking water sources only indicates that the water is piped and connects people to a supply system, making it accessible. The variable sets the criteria that the drinking water source is protected against external contamination, increasing the likelihood of high quality drinking water (WHO & UNICEF, 2021). Nevertheless, there is no further clarification on what measures are made to guarantee the quality of the drinking water. Piped drinking water sources can arguably be a measure of uncertain quality because of this. Additionally, out of the 12 indicators for water, this best represents a source that is universally accessible but with an ambiguous indication for quality. It is worth noting that the measure is therefore set as *undetermined quality* rather than *low quality*, but that the measure is still problematic and therefore coloured in grey in model 3.1.

Unimproved drinking water is the percentage or the population using unprotected sources of drinking water. These sources are not protected against contamination naturally by their construction and lack of active protection (WHO & UNICEF, 2021). Therefore, they are especially prone to bacterial contamination, and include sources such as unprotected wells or surface water.

3.5.3. Controlled variables

The variable *year* shows which year the data was collected for. Due to there being multiple observations for each country, the variable has been dummy-coded, making each year its own dummy variable. For each year, the year has the value of 1, while the remaining years are set as reference. This is done in order to help explain the variance in the dependent variable, with respect to the reference years. There are 222 observations for each year, except for 2018 and 2019 when there were 221 and 220 observations, respectively.

GDP per capita is a continuous variable, measuring income per capita. High GDP per capita is normally used to indicate economic prosperity and can therefore be used to measure economic development. Because of the high standard deviation, the variable is categorized into the three categories high income, middle income, and low income. The categorization is based on the mean threshold for income levels for the years 2014-2019, as defined by the World Bank Atlas Method (WB, b). For simplicity, upper middle income and lower middle income has been combined for a joint middle income category.

Minimum acceptable diet is a continuous variable measuring the percentage of children between 6-23 months old that consumed the minimum acceptable diet the previous day. Although the variable measures children, it includes breastfeeding and would therefore indirectly indicate the nutritional intake of the mother as well. The minimum acceptable diet includes both the frequency of meals as well as the diversity of their diet. The variable is from the Child Nutrition dataset by UNICEF. Minimum acceptable diet can be considered as a measure for the available nutrition in a country. A low measure would imply that there is not sufficient nutrition to cover the nutritional needs of the population.

Educational attainment is a continuous variable that measures what percentage of the population ages 25 and older that has completed primary education. Because the variable measures completed primary education in percentage, it is not subject to categorization and will remain a percentage. The objective of using this variable is to give an indication of literacy in the adult population.

Death rate is a measure for the number of deaths that occurs per 1 000 people per year. The variable is included because it gives an indication of health in the population. A low death rate suggests better health within the population than what is suggested in a population with a high death rate.

Life expectancy is the life expectancy within a country and is also a mortality measure. The variable indicates how many years a newborn would live if the current patterns in living standards and mortality at birth were consistent throughout its lifetime. The variable is included because it is an indicator of the overall health status in a country. A low life expectancy could for example indicate high infant mortality and a generally low survival rate when facing health issues.

Variables	%	mean	S.D.	Min.	Max.
Stunting		23.28	12.88505	1.20	56.60
Safely managed drinking water		72.906	29.87488	5.565	100.00
Limited drinking water		3.9062	6.52942	0.00	37.4973
Piped drinking water		74.013	27.825	4.268	100.00
Unimproved drinking water		5.2821	7.84984	0.00	36.2464
GDP per capita		17827.2	26579.69	228.2	189487.1
GDP per capita: categorized		2.245	0.64635	1	3
High income	36.137%				
Middle income	52.243%				
Low income	11.618%				
Minimum acceptable diet		26.19	19.24482	1.90	82.40
Educational attainment		83.35	20.76849	10.93	100.00
Death rate		7.581	2.62741	1.127	16.433
Life expectancy		72.55	7.58908	49.89	85.08
N = 1329*					

Table 3.2: Descriptive statistics for all variables. *Total number of observations.

3.6. Prerequisites

The data that is used in this paper consists of panel data, or longitudinal data. The data consist of one observation for each country between the years 2014-2019. Thus, there are six observations for each country. The data for each country will naturally be correlated on some level and the effect will consequently be overestimated if one conducts a regression analysis without any preconditions. To prevent a misleadingly low standard error and an unreliable result, each year and country is dummy-coded through clustering. Clustering identifies smaller groups within the data, which in this case can be identified both as year and country. Year and country will therefore be taken into consideration in the analysis. There is a varying degree of data coverage for the datasets, causing observations with missing values for the relevant variables to be excluded. The benefit of clustering may therefore be lost in analyses where many observations are excluded. However, clustering is applied to prevent that *year* and *country* cause unreliable results, but do not affect the analyses in the event that they do not cause unreliable results. Clustering the data is a preventative measure, only affecting the result in the event that there is a strong effect within the clusters.

Before conducting the analyses, the independent variables and controlled variables have been tested for collinearity. It is evident that the level of collinearity generally is high. This was expected for the independent variables and a risk with *model 3.1.*, where two variables automatically overlap for each categorization such as accessible drinking water. The controlled variables are all measures reflecting the level of development for a country and collinearity between these is not surprising either. As previously literature has noted, one cannot simply remove collinear variables with the expectation of a more accurate result (Freckleton, 2011, p. 99). Thus, the collinear variables will be allowed due to the reason behind the collinearity being known and the variables are explained as different development measures.

A null model was generated before the analyses. The variables were added one by one, which identified problematic relationships. The categorized *GDP per capita* caused an unrealistically high coefficient for the constant. When it was compared to the non-categorized measure of *GDP per capita*, the variable that was not categorized provided measures consistent with the other models. The categorized *GDP per capita* was therefore excluded in favor of the non-categorized version of the variable. When using *minimum acceptable diet* and the *GDP per capita* that was not categorized in the same model, the coefficient and standard error of the constant was abnormally high. A similar problem occurred when both *GDP per capita* and *life expectancy* were included in the same model. The constant shows the predicted Y-value when all X-values are zero. Although adjusting all X-values to zero is virtually impossible in itself, it is not a real possibility that the constant value of stunting would be >100. Due to this, *GDP per capita* was excluded from the model and a second model was made with *GDP per capita*, excluding *minimum acceptable diet* and *life expectancy*.

3.7. Analysis approach

Linear regression and multiple linear regression will be used for the analysis. Multiple linear regression is an extension of linear regression, which predicts the outcome of one dependent variable based on the effect of one independent variable, or several independent variables in the case of multiple linear regression.

(3)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$$

The Y represents the dependent variable stunting. Stunting is predicted by the explanatory variables X. The explanatory variables are the independent variables used to explain the variation in stunting. β_0 is the constant which is the value of the dependent variable stunting when all of the other variables that are included are set to zero. The beta for each explanatory variable is the slope coefficient that indicates the expected change in stunting (Y) for each unit increase in the explanatory variable (X). The residuals are included as ϵ.

(4)

$$Stunting = \beta_0 + \beta_1(safely\ managed\ drinking\ water) + \beta_2(piped\ drinking\ water) + \cdots + \beta_9(life\ expectancy) + \epsilon$$

For visualizing the effect of the different types of drinking water on stunting, a simple linear regression is used. Here, one independent variable is used to explain the variation in stunting. Linear regression is also referred to as ordinarily least squares (OLS), and the goal is to make a regression line with the least distance to each measure. The formula becomes a simple version of the formulas presented above.

(5)

$$Y = \beta_0 + \beta X + \epsilon$$

3.7.1. Software

The initial adjustments were made in Notepad++ (version 8.2.1, 32-bit). This was used in order to tidy the datasets, removing observational comments and assemble variables more efficiently.

The analysis and figures have been made in Rstudio (version 2022.02.2+485) with R version 4.0.2. The script including all functions for the analysis and explanations is included in annex 1.

4. Results

This chapter presents the analytical findings. The chapter starts by defining a level of significance and its implications for the results. Following this, the multiple linear regression for hypotheses H1-H5 will be presented along with a brief analysis of what model is the best fit. After this, the results of the multiple linear regression will be presented, first with a focus on the effect of accessibility of drinking water on stunting followed by the effect of quality of drinking water on stunting. Next, the multiple linear regression with interaction terms is presented and explained. Lastly, an effect plot is presented as a visual aid for detecting a pattern.

4.1. Level of significance

It is important to define a level of significance because it determines at what percentage the null hypothesis can be rejected with a high level of confidence. The level of significance is in some degree arbitrary, and we are at liberty to choose at what threshold we consider a value to be significant. Normally in the field of social science, the level of significance is defined as "$p < 0.05$" (Wasserstein, Schirm, & Lazar, 2019). When the p-value is below 0.05 it means that there is a 5% chance that the effect is random, and the null hypothesis can be rejected. The threshold of 0.05 has received criticism for being too accepting of effects, and an alternative level of significance of 0.005 has been proposed (Wasserstein et al., 2019). While the argument that 0.05 is too accepting of effects that may not exist is relevant, a level of significance of 0.005 may exclude effects that do exist. Because the aim is to determine if there is a relationship between drinking water and stunting it is beneficial to keep the level of significance as "$p < 0.05$" to prevent any wrongful exclusions of effects. Any levels of significance below this measure will further increase confidence in the existence of the effect.

4.2. The effect of accessibility of drinking water on stunting

Table 4 below presents the results from the multiple linear regression. The table will be used for hypotheses H1-H5, thus all comments in chapters 4.2. and 4.3. refer to the models within table 4.1.

As an initial model to test the independent variables, model 1 only includes the independent variables of drinking water sources. Next, models 2 and 3 include the controlled variables. As previously discussed in chapter 3.6., the two models differ in which of the controlled variables are included due to problematic results when they were included in the same model.

	Model 1	Model 2	Model 3
	Coefficient (Std. Error)	Coefficient (Std. Error)	Coefficient (Std. Error)
Constant	34.480 (3.192) ***	40.732 (116.293)	43.050 (5.776) ***
Safely managed drinking water	-0.114 (0.044) **	-0.121 (0.233)	0.088 (0.094)
Piped drinking water	-0.177 (0.036) ***	-0.225 (0.083) *	-0.171 (0.060) ***
Limited drinking water	0.029 (0.133)	-1.862 (1.732)	0.065 (0.298)
Unimproved drinking water	0.233 (0.122) *	0.320 (0.263)	0.513 (0.027) **
Minimum acceptable diet		-0.088 (0.315)	
GDP per capita			-0.000
Educational attainment		-0.250 (0.292)	-0.206 (0.121)
Death rate		1.725 (0.800)	-0.663 (0.451)
Life expectancy		0.147 (1.706)	
Adjusted R^2	0.6664	0.9117	0.8223
Obs. deleted	1217	1317	1300

N = 1329

*p<0.05, **p<0.01, ***p<0.001

Table 4.1: Multiple linear regression for stunting, predicted by the independent variables and the controlled variables.

By comparing the results of the multiple linear regression presented in the three models in table 4.1, it is possible to determine a model of best fit. The three models all have a relatively high explanatory power. It is evident from model 1 that the independent variables alone can explain more than half of the variance in stunting. The adjusted R^2 for model 1 indicates that over 66% of the variance in stunting is due to the independent variables of drinking water sources. Model 2 has the highest overall explanatory power, with its adjusted R^2 indicating that the variables included explain 91% of the variance in stunting. However, the standard error for the constant in model 2 is extremely high compared to its constant coefficient. The standard error indicates how far the measure is from the regression line which is why the high standard error is of concern. This suggests that the effects may be random. Considering the extremely high adjusted R^2, it is apparent that model 2 is problematic. This may be a result of eliminated observations with missing values, in which the remaining values are too few and have extreme values that do not give a realistic result. Lastly, the variables in model 3 explain 82% of the variance in stunting. Having a high degree of explanatory power and overall lower standard errors than model 2, the variables in model 3 are more consistent with the regression line and suggest that model 3 is the model of best fit.

The two measures of accessible drinking water are safely managed drinking water and piped drinking water while the two measures for non-accessible water are limited drinking water and unimproved drinking water. Piped drinking water is problematic because of its ambiguity in quality, but as a measure of accessibility it is accurate and can be used as such without any further comments. Using model 3 from *table 4.1*, it is evident that there is a significant difference in the effect that accessibility of drinking water has on stunting. Piped drinking water has a statistically significant negative effect on stunting. Thus, with higher levels of piped drinking water in the population it predicts a decrease in the level of stunting. The effect of piped drinking water is significant on a 0.001-level, causing a null hypothesis to be rejected with a high level of certainty. In other words, we can be confident that there is an effect of piped drinking water on stunting. The size of the effect however is moderate. Unimproved drinking water is statistically significant and has a has a positive and high effect on stunting, meaning that with increased unimproved drinking water the probability of stunting increases significantly. Relating this to their respective measures, accessible drinking water sources causes a decline in stunting while inaccessible drinking water sources causes an increase in the level of stunting. Safely managed drinking water and limited drinking water are not statistically significant. The lack of contrast between these two measures as seen in the other two variables of one positive and one negative coefficient is therefore not relevant for the discussion of accessibility. Based on this analysis, the likelihood of accepting the hypotheses H1 and H2 is strengthened.

None of the controlled variables are statistically significant. However, this does not mean that there are no effects, but that the probability of the effects being random is higher than 5%. It is more likely that the controlled variables are random, but this could also occur due to excluded datapoints not being able to support the effect. The controlled variables do contribute to the explanatory power of model 3. The other measures not significant. However, they do contribute to the explanatory power. Comparing model 3 to model 1, adding the controlled variables explains over 15% more of the variation in stunting. The controlled variables give estimations of nations' level of development, providing insight into basic conditions that may affect stunting. The variable GDP per capita has a negative effect indicating a decline in stunting, but this effect is minimal and not considered an effect at all. Educational attainment also indicates a decline in stunting. Death rate has a high effect that indicates a decline in stunting, but the high standard error supports the argument of it being random and not an actual effect.

4.3. The effect of drinking water quality on stunting

The two measures for high quality of drinking water sources are safely managed drinking water and limited drinking water, and the measures for undetermined quality are piped drinking water and unimproved drinking water. Undetermined quality does not necessarily mean low quality, but the likelihood for contamination is higher. As discussed, when including the variable of piped drinking water, the likelihood of it representing high quality is high but not certain due to it being a protected source. Thus, there may be some overlap with the measures for high quality. Throughout the models, there is inconsistency within the measures of high quality and undetermined quality. To prove that quality has an effect on stunting, the measures within each category must indicate a similar direction, both either indicating an increase or decrease in stunting.

In model 1, safely managed drinking water is statistically significant and indicates a decline in stunting. Limited drinking water is not statistically significant and indicates an increase in stunting. In model 2, both measures indicate a decrease in stunting although not statistically significant. In model 3, safely managed drinking water and limited drinking water indicate an increase in stunting, but the measures are not statistically significant, and the effect is marginal. The variables of piped drinking water and unimproved drinking water, as measures of undetermined quality, are both statistically significant. There is an inconsistency within this category due to piped drinking water indicating a decline in stunting while unimproved drinking water indicating an increase in stunting. This can be caused by the variable piped drinking water which may have an overlap with high quality drinking water.

Safely managed drinking water has a statistically significant negative effect on stunting in model 1. The result of model 1 is that countries with higher levels of safely managed drinking water have a lower probability of stunting, or a decrease in the level of stunting. After the controlled variables are added, safely managed drinking water loses its statistical significance. Compared to the other variables for drinking water, safely managed drinking water does not have an increase in its effect either from model 1 to model 3. Safely managed drinking water has a limited effect on stunting and is not considered statistically significant in model 3. It is not possible to confirm the hypotheses H4 or H5 because the likelihood of the effect being random is higher than 5%, which is the set threshold for statistical significance. Additionally, there are other measures that are more statistically significant with a proven correlation after the standards set in this paper.

4.4. Examining the effect of food and water on stunting

Table 4.2 below show the multiple linear regression with interaction terms. The interaction terms have been added to explore if the variables for drinking water and the measure of available food are dependent on each other. If any of the interaction terms are significant, it would indicate that an increase in one variable give the other variable more effect.

For this analysis, no control variables are added other than *minimum acceptable diet* which is included in the interaction term. None of the remaining controlled variables are significant, thus there is no loss of significant effects by not adding these. Any effects of the interaction terms may be overestimated by not including controlled variables. Nevertheless, overestimating an effect establishes a relationship and give grounds to further examine the relationships which is productive.

	Model 4	Model 5	Model 6
	Coefficient (Std. Error)	Coefficient (Std. Error)	Coefficient (Std. Error)
Constant	30.115 (11.117) ***	40.427 (4.509) ***	29.446 (7.205) ***
Safely managed drinking water	-0.015 (0.153)		-0.289 (0.116)
Piped drinking water	-0.263 (0.104) **	-0.290 (0.064) ***	
Limited drinking water	0.765 (0.447) *		0.774 (0.448)
Unimproved drinking water	-0.017 (0.381)	-0.069 (0.186)	
Minimum acceptable diet	0.127 (0.380)	-0.401 (0.160) **	0.269 (0.247) **
Safely managed drinking water * minimum acceptable diet	-0.004 (0.003)		-0.002 (0.003)
Piped drinking water * minimum acceptable diet	0.003 (0.003)	0.003 (0.002) *	
Limited drinking water * minimum acceptable diet	-0.032 (0.020)		-0.003 (0.020) *
Unimproved drinking water * minimum acceptable diet	0.017 (0.018)	0.025 (0.010) **	
Adjusted R^2	0.5571	0.528	0.4716
Excluded obs.	1264	1225	1264

N = 1329

*p<0.05, **p<0.01, ***p<0.001

Table 4.2: Multiple linear regression for stunting, predicted by the independent variables and interaction terms with minimum acceptable diet.

Model 4 is an initial model including all the relevant variables. The explanatory power is relatively high, and from the adjusted R^2 it is evident that the variables included in model 1 explain over half of the variance in stunting. In model 4 there are only two statistically significant measures, namely piped drinking water and limited drinking water. Additionally, the effect is substantial. Piped drinking water indicated a decrease in stunting by having a negative effect, while limited drinking water indicates an increase in stunting by having a positive effect on the dependent variable. The interaction terms in model 4 are statistically insignificant and have a minimal effect on stunting.

By comparing the results in model 5 and model 6, model 5 has the higher explanatory power. The adjusted R^2 shows that model 5 explains over half of the variance in stunting while model 6 explains just under half of the same variance. The measure of standard error in model 5 are overall lower than the standard errors in model 6, suggesting that the datapoints of the variables in model 5 are closer to the regression line and follow that pattern. It is not of great importance to determine a model of best fit in table 4.2 because they include alternate independent variables. Model 5 includes the independent variables for undetermined quality drinking water while model 6 includes the independent variables for high quality drinking water. The interaction terms that are added follow this pattern. The only variable that is included in both models is *minimum acceptable diet*. This variable is interesting because it is statistically significant in both models but differ in the effect it has on stunting. Because model 5 has a higher explanatory power, it is reasonable to believe that the variables and their measures in this model are more appropriate for explaining the variance in stunting.

In models 5 and 6, piped drinking water is the only independent variable with statistical significance and indicates a decline in stunting. Limited drinking water in model 6 has a high effect but paired with the relatively high standard error it supports it being random with no actual effect, which is apparent by the variable not being significant as well. In model 5, *minimum acceptable diet* is statistically significant and indicates a decline in stunting with a high effect. In model 6, the same variable is statistically significant and indicates in increase in stunting. The effect is not as high as in model 5, and with a higher standard error. However, as previously stated, the variable *minimum acceptable diet* in model 5 is a more appropriate measure for the effect on stunting.

The effects of the interaction terms are minimal. In models 5 and 6, the interaction terms of piped drinking water and minimum acceptable diet, limited drinking water and minimum acceptable diet, and unimproved drinking water and minimum acceptable diet are statistically significant at least on a 0.05-level. The variance in stunting is therefore likely to be affected by the interaction of the two variables in each of the significant interaction terms. However, because of the marginal size of the effects it must be considered if they prove an actual effect or not. For unimproved drinking water, the interaction term with minimum acceptable diet can be considered having an actual effect due to the size of the effect being higher, the standard error significantly lower, and the interaction term is statistically significant. For the remaining interaction terms, the effects are smaller than the effects of each of the variables alone. Comparing the models 5 and 6 to model 4, the size of the effect is considerably smaller. Thus, the interaction term of unimproved drinking water and minimum acceptable diet has a greater effect on stunting than unimproved drinking water alone. For the other interaction terms this is not true.

The coefficients for the interaction terms hold other information as well. The coefficients for undetermined quality in model 5 are positive, suggesting a mutually reinforcing relationship in the interaction terms. Thus, an increase in piped drinking water causes a higher positive effect of minimum acceptable diet on stunting, which indicates an increase in stunting. The positive coefficient can also indicate a decrease in piped drinking water that causes a higher negative effect of minimum acceptable diet on stunting, which is a decrease in stunting. It is however not implied which of the variables affect the other. The coefficients for high quality in model 6 are negative, suggesting that there is a contrast in the effect of the variables in the interaction terms. Thus, an increase in one variable causes the other variable to have a negative effect on stunting, indicating a decrease in stunting. It is also possible that the decrease of one variable in the interaction term causes the other variable to have a positive effect on stunting, indicating an increase in stunting. For these measures it is also difficult determining which of the variables has the effect on stunting as a result of the effect from the first variable.

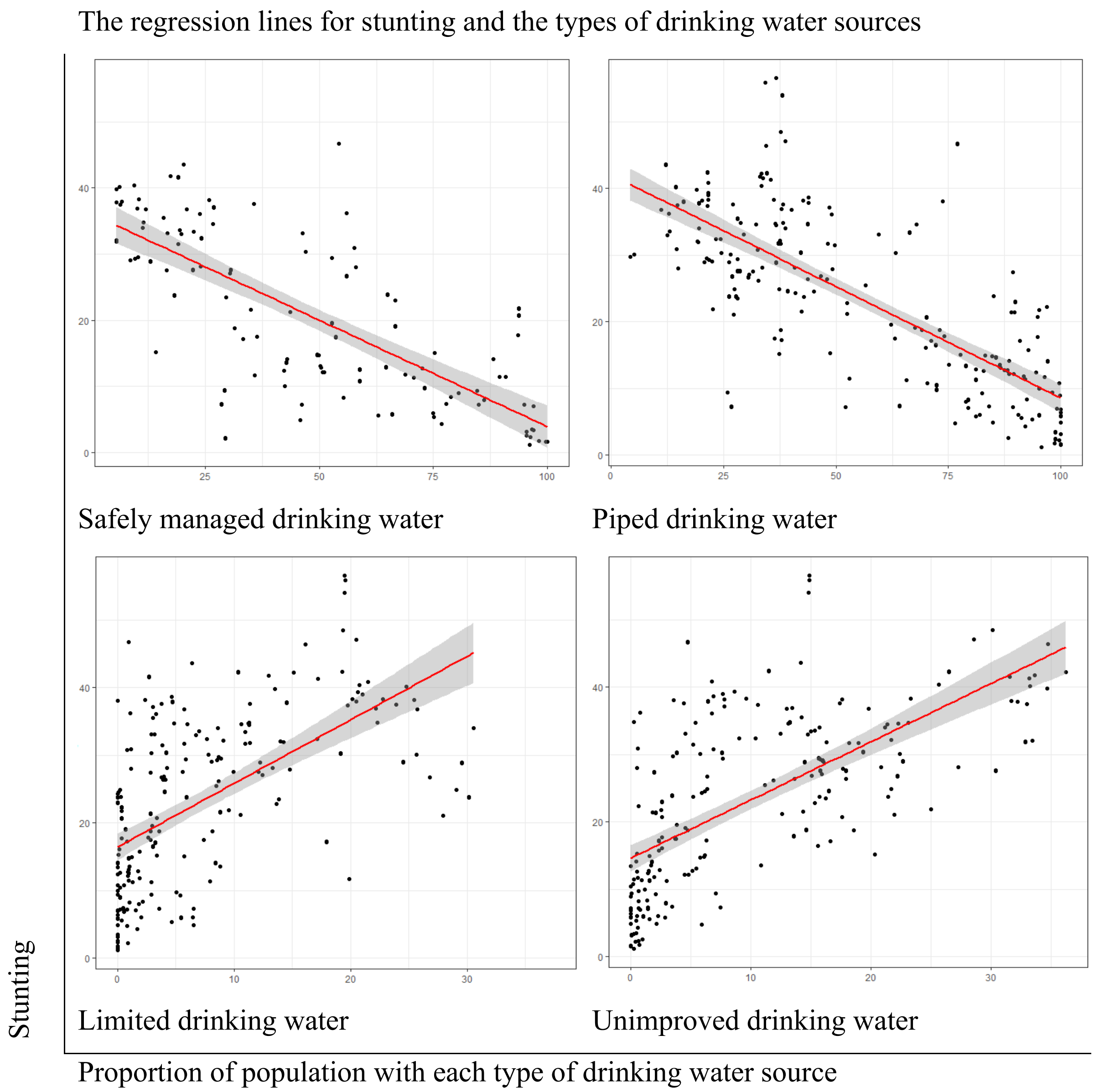


Figure 4.1: The regression lines from linear regression (OLS) for stunting and the types of drinking water sources.

In addition to the analysis of tables 4.1 and 4.2, it is useful to generate a visualization to determine any patterns between the independent variables and their effect on stunting. Figure 4.1 shows the regression lines from linear regression for each type of drinking water source and their effect on stunting. The linear regression is the effect of one independent variable on one dependent variable and is therefore not controlled for other variables. The plots are presented in a similar way as the variables were presented in table 3.1, where the two measures for accessible drinking water sources are above the two measures for inaccessible drinking water sources. The left column represents drinking water sources of high quality and the columns to the right represent drinking water sources of undetermined quality. From figure 4.1 it is evident that there is a consistent pattern in the measures of accessible drinking water and a different consistent pattern for the measures of inaccessible drinking water. The regression lines clearly show the direction of the relationships. Accessible drinking water indicates a decrease in the level of stunting within a population, while inaccessible drinking water indicates an increase in stunting. The plots for the measures of accessible drinking water show a straight regression line. The plots for limited drinking water and unimproved drinking water are more ambiguous. Unimproved drinking water has previously proved to be statistically significant, but from the visualization in figure 4.1 it is not clear if the regression line should remain a straight line of it has a more curved regression line. If the plots indicate a curved line, it provides an explanation as to why the measures of inaccessible drinking water are considered random in the analysis, and the possibility for examine the relationships with a new type of regression analysis.

5. Discussion

In this chapter the results from the analysis will be discussed. The chapter begins with a commentary on the models presented in chapter 4 and the implications of their contents. Following this is a discussion of each of the research questions, based on the results. The discussion and conclusions of the hypotheses will then be summarized. Lastly, the critique of the study will discuss areas of improvements.

5.1. Commentary on models

The controlled variables are all variables used to measure the level of development. Similarly, the independent variables of drinking water sources can reflect aspects of the society in which the populations live. Water access and quality can for example indirectly reveal if the infrastructure is prioritized, such as facilitating piped and treated drinking water. Access to water can also imply strategic city planning, that the population has sufficient resources to purchase drinking water or that it naturally exists in a quantity that fulfills the needs of the population.

There is a difference in the number of observations across the datasets. During the regression analysis, observations with missing data for the relevant variables have been excluded as previously discussed in chapter 3.6. The exclusion of variables affects the models equal to having fewer data points and consequently less information. The results in the models may therefore be exaggerated or underestimated, and gives results based on the remaining observations. The models indicate directions of relationships or if any effects are significant but would provide more accurate results with a higher number of observations.

Table 4.1 has a notably high adjusted R^2 for all models. It is possible that this is due to the effects being exaggerated by missing data points as described above. However, the independent variables and controlled variables all reflect aspects of development and standards of living. These are the basic conditions for the development of an individual, and therefore have a high significance for the probability of stunting. It is therefore feasible that the explanatory power is high.

5.2. Accessible drinking water as a determinant for stunting

The effect of accessibility of drinking water on stunting has been tested by comparing the two measures of accessibility (safely managed drinking water and piped drinking water) and inaccessibility (limited drinking water and unimproved drinking water).

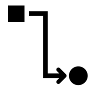
	High quality	Undetermined quality
Accessible	Safely managed drinking water 	Piped drinking water sources 
Not accessible	Limited drinking water services 	Unimproved drinking water source 

Table 5.1: Measures of accessible and inaccessible drinking water sources.

In the analysis, piped drinking water has remained statistically significant in all models, proving that the effect of piped drinking water on stunting is not a random effect. The variable is consistently negative, implying that higher levels of piped drinking water reduce the level of stunting. This supports the assumption in hypothesis 2, that the independent variable piped drinking water significantly reduces the probability of stunting. It is problematic to conclusively verify the relationship without further examination, but the findings strongly suggest that hypothesis 2 is an accurate interpretation of the relationship.

Safely managed drinking water also indicates a decline in stunting in all models except for model 3 in table 4.1. In this model, the variable has a small effect indicating an increase in stunting. Safely managed drinking water is only statistically significant in model 1 in table 4.1, where the variable indicates a decline in stunting. The findings from the analysis are therefore consistent with hypothesis 1, that accessible drinking water suggest a decline in stunting. In contrast, the variables for inaccessible drinking water indicate an increase in stunting. In table 4.1, unimproved drinking water is statistically significant in models 1 and 3, with a high effect that shows an increase in stunting. In table 4.2, limited drinking water has a high effect on stunting that is statistically significant. Comparing the effect of the variables for inaccessible drinking water to that of piped drinking water, there is a considerable difference. The importance of the effect of accessible drinking water would not be apparent without a comparison. Thus, the contrasts between accessible and inaccessible drinking water further support hypothesis 1.

Inaccessible drinking water sources are problematic for the nutritional well-being and consequently the probability of stunting. Considering the theories of the interdependence of food and water, drinking water is a part of the total nutritional status (Choudhary et al., 2021; Popkin et al., 2010). Inaccessible drinking water sources imply that there are barriers that make it more difficult to collect drinking water. This poses a risk to the nutritional well-being because the collection of water requires a level of energy that is not universally expendable. Collecting drinking water from inaccessible sources requires a higher energy output, and there are multiple factors that can affect this. First, inaccessible sources may require more energy due to the sources being further away from the home or hard to physically reach. Approaching the drinking water source and then returning with the collected water can be considered physical work that requires more energy compared to what is required for collecting tap water within the home. Second, Inaccessible water requires time, which for the variable limited drinking water is >30 minutes for a round trip (UNICEF, b). Thus, inaccessible sources can affect work capacity both by reducing the remaining energy for work as well as limiting the amount of time a person can work. Referring to the physical capacity circularity, inaccessible water prevents a positive circle of increased nutrition, work capacity, and income. Lastly, when basic needs are not being fulfilled it is natural that the individual strategize how to fulfill their needs. These concerns regarding drinking water equal cognitive energy expenditures that require a higher level of energy (Popkin et al., 2010). Making collection strategies, organizing a daily routine to include the collection of water, or searching for alternative drinking water sources all require energy to do and this is why cognitive energy expenditure is higher with inaccessible drinking water.

Accessible drinking water diminishes the energy output required for the collection of drinking water. The surplus energy is thus disposable for other activities, such as work capacity or other measures to attain more nutrition. In accordance with the physical capacity curve, surplus energy is used for labor to increase income and subsequently provide nutrition. Accessible drinking water can therefore also indirectly affect the level of stunting by facilitating increased income and nutrition intake.

5.3. Quality of drinking water as a determinant for stunting

Similar to the measures of accessibility, the effect of the quality of drinking water on stunting has been tested by comparing the effects of the two measures of high quality (safely managed drinking water and limited drinking water) and the two measures of undetermined quality (piped drinking water and unimproved drinking water). Undetermined quality is selected as a contrast to high quality drinking water sources due to the absence of a standard for levels of contamination. It is not a perfect opposite of high quality and can therefore overlap measures for high quality drinking water, as discussed in chapter 5.1.

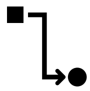
	High quality	**Undetermined quality**
Accessible	Safely managed drinking water 	Piped drinking water sources 
Not accessible	Limited drinking water services 	Unimproved drinking water source 

Table 5.2: Measures of high quality and undetermined quality of drinking water sources.

The measures for high quality drinking water are conflicting. Safely managed drinking water is significant in model 1, before adding the controlled variables, indicating a reduction in the level of stunting. Limited drinking water is significant in model 4, where the effect indicates a considerable increase in the level of stunting, but with a high standard error. The conflicting results and the lack of consistent statistical significance of high quality drinking water does not support the existence of a common effect for high quality drinking water. Furthermore, this does not provide a solid reference for comparing results with the effects of undetermined quality drinking water. Thus, there is no clear divide between the results of the effects of high quality drinking water versus the effects of undetermined quality of drinking water on the level of stunting. As a result of this interpretation, hypothesis 3 is rejected and high quality of drinking water does not cause a decline in stunting.

When the interaction terms are added, an interesting pattern in high quality drinking water becomes apparent. Model 6 reveals that the interaction term of limited drinking water and minimum acceptable diet is statistically significant and negative. This means that as the effect of limited drinking water on stunting increases, the effect of minimum acceptable diet reduces, or vice versa. This can be understood as quality being of importance. Better quality of drinking water weakens the effect of minimum acceptable diet on stunting, which may be caused by less contamination and water-borne diseases interfering with the body absorbing the available nutrition. This argument is partially supported by the notion of drinking water being the most important part of a diet, more so than food that actually contain nutrients (Popkin et al., 2010). On the other hand, as a higher proportion of the population obtain the minimum acceptable diet, The quality of drinking water loses its effect. A higher proportion of the population where the minimum acceptable diet is met is telling on the populations level of development. Thus, limited drinking water may lose its effect because the need for quality drinking water is already met or due to the population having sufficient nutritional intake for constant replenishment.

Ultimately, considering the contrasting effects of good and undetermined quality drinking water on stunting individually, the conclusion is a rejection of hypothesis 3. Factoring in the interaction terms, they are more telling of the interconnectedness of the availability of nutrition and drinking water and its effect on stunting. Although the interaction terms provide added complexity to the issue of determinants for stunting, they do not support hypothesis 3.

Safely managed drinking water is the aim in the SDG 6.1 (WHO & UNICEF, 2021, p. 5). It appears to be a kind of golden standard for accessible and high quality drinking water. The initial assumption was therefore that safely managed drinking water would have the most profound effect on stunting, as well as indicating a decline in stunting as safely managed drinking water increased. However, the variable safely managed drinking water is significant only in model 1, implying that the effect shown in model 1 is owed to the controlled variables added in later models. Both in effect size and significance, safely managed drinking water is not the main determinant for stunting, thus rejecting hypotheses 4 and 5.

5.4. Inclusion of drinking water into the theory of the physical capacity curve

In order to answer research question 3, the discussion relies heavily on both the analysis as well as the theoretical framework. The results from the analysis are examined using the perspective of the theories presented in chapter 2.

In the theories of interdependence of food and water, it is stated that drinking water is required for nutritional uptake (Choudhary et al., 2021, p. 171). It implies that the body cannot absorb nutrients from food properly without the presence of water, supporting the inclusion of drinking water in the total nutritional status. Interaction terms were generated for minimum acceptable diet and each of the variables for drinking water source, aiming to test if the interdependence of food and water has a stronger effect on stunting than nutrition and drinking water separately. In model 4, the interaction terms are not significant. There are four interaction terms in total, of which all include the variable minimum acceptable diet as a measure of available nutrition. The lack of significance may be due to the fact that the same variable is used for all four interaction terms, causing any effect to be missed. However, the interaction terms are separated in models 5 and 6 arbitrarily based on quality of drinking water, causing significant effects to appear.

In model 5, the interaction term consisting of piped drinking water and minimum acceptable diet is statistically significant. This result in itself is supported by the theories that there exists an interconnectedness between nutrition and drinking water, and that one is dependent on the other to sustain the body (Popkin et al., 2010). The positive coefficient for the interaction term of piped drinking water and minimum acceptable diet suggests that an increase in one of the independent variables causes the other independent variable to affect an increase in the level of stunting. Conversely, this proves that if levels of either available nutrition or piped drinking water decreases it affects the other independent variable negatively, indicating a decrease in the level of stunting.

Considering the effect of accessible drinking water, this effect proves contrary to the literature stating that drinking water is needed to efficiently absorb nutrition which in turn is required to prevent stunting from as early as conception (Black et al., 2013; Choudhary et al., 2021). The decrease in the level of stunting was therefore expected to be due to the existence of available nutrition as well as drinking water, not the lack thereof. The effects of the interaction terms including piped drinking water and unimproved drinking water are both positive, showing no contrast between the measures of accessible and inaccessible drinking water, respectively. Piped drinking water is a measure for accessibility as well as undetermined quality. The effect of the interaction term can therefore be due to a higher probability of diseases, such as diarrheal diseases where there is a lack of nutritional uptake into the body caused by the disease. Consequently, the minimum accepted diet shows an increase in stunting due to the loss of energy input, or a nutritional and energy deficit.

The effect size of the interaction terms including the measures of accessible drinking water is not greater than the effect size of accessible drinking water individually. The measures of accessible drinking water individually show a consistent effect on stunting, proving a decrease in the level of stunting compared to the effect of inaccessible drinking water. Although the interaction terms for accessible drinking water and minimum acceptable diet are significant, the effect size is small and lack a theoretical foundation. Thus, the conclusion is that hypothesis 6 is rejected.

Model 6 include the interaction terms that include the measures for high quality drinking water. The interaction terms both have a negative coefficient, proving for example that an increase in one of the independent variables affect the other independent variable to indicate a decrease in the level of stunting. It is reasonable to believe that the availability of nutrition is dependent on high quality drinking water to lower the probability of stunting in children. Drinking water is a basic need that is required for nutritional uptake and consequently energy input. Therefore, it is rationalized as a conditional variable that the independent variable minimum acceptable diet is dependent on in the interaction term.

The availability of nutrition, as shown by the proxy minimum acceptable diet, is a main factor in several concepts. For example, malnutrition and undernutrition is widely accepted as the determinant for stunting (WHO, 2014). The lack of nutrition or the lack of nutritional uptake in the body is thus recognized as a causing factor for stunting. The presence of low quality drinking water may increase the risk of diseases such as diarrhoea that prevent optimal nutritional uptake (Jain, 2012; Todaro & Smith, 2012). This makes drinking water quality a condition for the direct effect of nutrition on the probability of stunting. Additionally, the intake of nutrition is regarded as energy input in the theory of energy balance, implying that nutrition is the main cause for the continuous energy cycle in the body. However, similarly to the point above, low quality may hinder this by preventing optimal nutritional intake. Lastly, nutrition is explicitly mentioned as the determinant for work capacity in the physical capacity curve (Dasgupta & Ray, 1986). The examples above support the argument of nutrition being the direct link, while drinking water is the conditional variable in the interaction terms.

When examining the measures of high quality drinking water individually, safely managed drinking water is statistically significant in model 1. Outside of model 1, the measures for high quality drinking water do not show a significant effect. On the other hand, the measures for undetermined quality drinking water are statistically significant in model 3. This result is supported by previous literature that point out that low quality of drinking water could make an individual more susceptible to infections and diseases (Haq et al., 2007; Todaro & Smith, 2012). Low quality of drinking water increases the likelihood of negative health outcomes such as diarrhoea, skin irritation and physical abnormalities, which in turn can affect the probability of stunting (Hunter et al., 2010, pp. 2-3). The measures of undetermined quality can be interpreted as a reference for further assumptions. However, it would be wrong to use these findings to conclude on the results for the effect of high quality drinking water on the level of stunting. Although it is unlikely to be practical, the effect of undetermined quality of water should remain of interest for further studies.

There is a strong indication of an effect between quality of drinking water and stunting, based on previous literature and the theoretical framework. The discussion has supported this, further suggesting an interconnectedness between nutrition and drinking water that is dependent on high quality of the drinking water. The interaction term consisting of limited drinking water and minimum acceptable diet shows a significant effect, although the effect size is small. However, compared to the measures of high quality drinking water individually, the interaction term shows a clear effect that is not random. Therefore, it is possible to conclude that hypothesis 7 is strengthened.

An interesting observation is the difference in how the drinking water variables individually affect the level of stunting versus how the interaction terms affect the level of stunting. Individually, the independent variables for drinking water show a pattern based on accessibility. This conclusion is further supported by the visualizations in figure 4.2. The trend relates to the theory of energy balance, and can be interpreted as more accessible drinking water lowers the energy output by not requiring additional energy for drinking water collection. Thus, more energy remains to do productive work, such as income-generating work or amassing food. Although accessible drinking water do not give a surplus of energy, it allows the remaining amount of energy to be used on personal or collective development, including nutritional well-being.

The interaction terms reveal a more complex dynamic between the available nutrition and drinking water, and their joint effect on the level of stunting. The significance of the interaction terms in models 5 and 6 may be the result of the exclusion of confounders, making the effect more limited and precise. As stated in hypotheses 3, 4, 5 and 7 regarding the quality of drinking water, the focus is on high quality drinking water. Therefore, the interaction terms including safely managed drinking water and limited drinking water are the main points of discussion. A common pattern for the interaction terms including the measures for high quality drinking water is not evident due to the interaction term including limited drinking water having the only significant effect on stunting. However, the joint effect of limited drinking water and minimum acceptable diet on the level of stunting can be interpreted as high quality drinking water indirectly showing a decrease in the level of stunting. The availability of nutrition, or lack thereof, is acknowledged as a determinant for stunting (WHO, 2014). Thus, it is expected that nutrition has the direct effect on the level of stunting, being dependent on the quality of water.

Another interesting point is that stunting is a hindered development, which may have negative consequences on work capacity. As the quality of drinking water is a condition for the effect of available nutrition on the level of stunting, it proves that high quality drinking water should be considered in further discussions about the effects of nutrition. Rather than theorizing on the effect of purely nutrition, high quality drinking water can be included in the concept of a minimum required diet. This rationale is interesting in the context of including drinking water into the physical capacity curve and is worth further examination.

Hypothesis 8 is open for interpretation, inviting and encouraging discussion. The measure for high quality as well as accessible drinking water is safely managed drinking water, which has not demonstrated a significant pattern in its effect on the level of stunting. However, there is a pattern in the individual independent variables, showing a direct link between the access of drinking water and stunting. Additionally, the interaction term for limited drinking water and minimum acceptable diet supports the importance of quality of drinking water in the effect of minimum acceptable drinking water on the level of stunting. The concept of pattern in hypothesis 8 is an explorative concept. It relates to the independent variables for high quality and accessible drinking water having a distinct effect on the level of stunting. The variables piped drinking water and undetermined quality drinking water individually, in addition to limited drinking water in the interaction term, have a specific and significant effect on stunting. Throughout the analysis, these effects have been discussed and supported by the theoretical framework. Consequently, the conclusion for hypothesis 8 is ultimately strengthened.

The research question of whether the relationship between drinking water and stunting is strong enough to support the inclusion of drinking water into the physical capacity curve cannot be answered solely with this thesis. However, there are several interesting points as explained throughout the discussion that proves the value of drinking water in theories including nutrition. The discourse of the effect of nutrition on any bodily abilities, functions or development can be considered weaker without the inclusion of drinking water. Finally, this thesis does not provide a conclusive answer to the research question 3 of whether the relationship between drinking water and stunting is strong enough to support the inclusion of drinking water into the physical capacity curve. Nevertheless, this thesis contributes to enlighten the issue of drinking water in relation to nutrition and urging further analysis on the topic.

5.5. Summary of results

The conclusions of the hypotheses are presented in tables 5.3, 5.4, and 5.5 below, with a short explanation included for each hypothesis.

Research question 1: Is there an association between accessibility of drinking water and stunting?		
Hypothesis	Conclusion	Explanation
1. Accessible drinking water causes a decline in stunting	Strengthened	The measure of accessible drinking water, piped drinking water, indicates a decline in the level of stunting.
2. Piped drinking water significantly reduces the probability of stunting	Strengthened	Piped drinking water is consistently a statistically significant measure that indicates a decline in the level of stunting.

Table 5.3: A summary of the findings for the hypotheses for research question 1.

Research question 2: Is there an association between quality of drinking water and stunting?		
Hypothesis	Conclusion	Explanation
3. High quality of drinking water causes a decline in stunting	Rejected	There is no significant effect of high quality drinking water in the model of best fit.
4. Safely managed drinking water significantly reduces the probability of stunting	Rejected	Safely managed drinking water loses its significance when adding controlled variables, rejecting the significance of the variable.
5. Safely managed drinking water has the most significant effect on stunting	Rejected	Safely managed drinking water is not the most significant measure, nor does the variable have the greatest effect on the level of stunting.

Table 5.4: A summary of the findings for the hypotheses for research question 2.

Research question 3: Is the relationship between drinking water and stunting strong enough to support inclusion of drinking water into the physical capacity curve?		
Hypothesis	Conclusion	Explanation
6. Accessible drinking water and available nutrition has a greater effect on stunting than accessible drinking water alone	Rejected	The effect size of the interaction term is smaller than the effect size of the measures of accessible drinking water individually.
7. High quality drinking water and available nutrition has a greater effect of stunting than high quality drinking water alone	Strengthened	The interaction term for limited drinking water is significant, justifying the importance of high quality drinking water on stunting. The variables for high qual drinking water individually show no significant effect.
8. There is a pattern between accessibility and quality of drinking water and stunting	Strengthened	Accessibility and quality of drinking water have separate but significant effects on stunting. Accessible drinking water indicates a direct effect on stunting. High quality drinking water is significant in the interaction term with minimum acceptable diet.

Table 5.5: A summary of the findings for the hypotheses for research question 3.

The conclusions made in the discussion are done in light of the hypotheses and based on the results presented in this thesis. Although the findings are justified, they may not give a definite conclusion. There always exists opportunities for further examination, which is especially true in this case. The attempt to modify a theory is not completed in one single analysis. The research in this thesis is a suggestion to further research the topic, as well as which effects can be expected. Additionally, technical aspects such as the exclusion of variables have presented some results that are not representative of the global population. Consequently, the indications of this research are initial assessments, that can show probable relationships that must be examined further to provide final, definite conclusions.

The results of the regression analysis show that the more accessible drinking water is, the more likely it is to observe a lower level of stunting. Model 3 has one of the highest explanatory powers of all the models and shows the contrast between accessible and inaccessible drinking water. Although the model also indicates the exclusion of many observations, the results give reason to believe that the effect of accessible drinking water on stunting exists. This is supported by the theory of energy balance, by that accessible drinking water prevents energy and time spent on collection of drinking water, which instead can be used in effective work. The accessibility of drinking water can also indicate better infrastructure, security in the supply of drinking water for the population, and generally a higher level of development (Hunter et al., 2010).

There is no evidence in the findings that suggest that quality of drinking water directly affects the level of stunting. However, the findings indicate that nutrition is dependent on the quality of drinking water for its effect on stunting. The effect of accessibility can also be connected to quality, as the lack of high quality drinking water may cause people to find drinking water from alternative, sometimes unsafe, sources (Edokpayi et al., 2018). Previous literature supports the indirect link between drinking water and stunting. For example, Todaro and Smith (2012, p. 489) explains that populations exposed to contaminated water more frequently experience diarrhoea, which in turn cause malnourishment and may affect the probability of stunting. Several other studies support this, connecting low quality drinking water to severe health risks affecting nutritional uptake, that in turn can affect the physical development (Choudhary et al., 2021; Edokpayi et al., 2018; Jain, 2012). In sum, inaccessible drinking water and drinking water of low quality may lead to diseases connected to nutrition. Consequently, proper uptake of nutrients is prevented, causing a state of malnourishment that increase the probability of stunting. Conversely, safely managed drinking water is linked to the alleviation of poverty, lower mortality rates, and enhanced productivity (Edokpayi et al., 2018; Haq et al., 2007; Jain, 2012). Thus, the findings here are consistent with the previous literature connecting drinking water indirectly to negative health effects such as stunting.

The aim of the interaction terms is to consider the interconnectedness of drinking water and nutrition. The results show that nutrition is dependent on drinking water, or vice versa, for its effect on stunting. A consequence of this is to consider changing the term of nutrition to that of diet to include a notion of the importance of drinking water. Through the discussion it has become evident that it is beneficial to consider drinking water because both describe basic needs that effect the development of the body. However, considering the excluded observations and the effect size of the interaction terms, the effect is not strong enough to conclude with an inclusion of drinking water into the theory of the physical capacity curve. Although the effect sizes of the interaction terms are low, their significance highlights the advantage of factoring in drinking water in the discussion of physical capacity in any theory. Ultimately, there exists an interconnectedness between nutrition and drinking water that jointly effects physical abilities.

5.6. Critique of the study

The independent variables of drinking water sources are not mutually exclusive variables. For piped drinking water and unimproved drinking water, the quality is uncertain (WHO & UNICEF, 2018). This implies that there is no guarantee that the water is free from contamination or safe to drink, and the variables were chosen because they are more likely to carry contamination. However, the variables do not exclude the possibility that the water from these sources can be of high quality and safe to drink. While it is possible to study the effect of high quality drinking water on stunting, providing a point of reference is problematic. Uncertain quality of drinking water is not the equivalent to have-not in regard to quality, therefore it does not show the contrasting effect to high quality drinking water. Comparing the effect of quality of drinking water on stunting is therefore challenging because there is no variable for drinking water that guarantees contaminated and unsafe drinking water.

There is a high correlation between the independent variables and the controlled variables. The study could therefore benefit from using a more advanced type of regression. Ridge regression is an alternative type of analysis approach that minimizes the effect of collinearity by adding a penalty term. Ridge regression requires complete observations. In the datasets used for this study, there are few complete observations to accurately predict effects in ridge regression. Producing estimates through automatic imputation is an alternative, but also problematic because it can have unintentional implications. It, therefore, requires a level of expertise to create automatic imputation for the data and use it ridge regression.

The literature in the field of social sciences on the interdependence of food and water is limited. This causes a high degree of assumptions and effects that are not adequately proven, such as the argument that water is the most important nutrient or the contribution of water output in resting metabolism through thermoregulation (Popkin et al., 2010, pp. 446-447). It is important to generate hypotheses that can be tested and contribute to proving or rejecting these assumptions. Some of the concepts used in the literature and in this study are ambiguous. The concept of nutrition was initially used for consumption of nutritional value but considering arguments for the value of drinking water it incorporates consumption for sustaining the human body, which includes drinking water. Work capacity is also an ambiguous concept with different possibilities for measurement. The literature presents different measures of work capacity, ranging from how many units a person can produce of a certain item to oxygen uptake while doing physical work (Dasgupta, 1997). The common denominator is at what rate an individual can convert energy into physical work, as defined in chapter 2.5. The assumptions and ambiguous terms are problematic if not clearly defined. Without a clear understanding, ambiguous concepts are problematic for the ability to generalize any findings and pose a risk for further research.

The lack of literature on the interdependence of food and water as well as on the physical capacity curve gives the existing literature greater power. There are few competing perspectives, suggesting that the literature that exists presents the solution. Comparing this to the literature regarding other topics of development, such as discourses of development or economic development, the topic of water is under-communicated. This further supports the argument above to provide clarifications that can be productive for further studies on the subject of the effect of water on nutritional status.

6. Conclusion

This chapter provides the final conclusions for each of the research questions. Following the conclusions are final comments to this study. Lastly, several suggestions for further research are presented.

6.1. The correlation between accessibility of drinking water and stunting

Research question 1 questions if there is a correlation between accessibility of drinking water and stunting. The research question is an initial step to determine if there are significant effects of drinking water on the level of stunting based on accessibility. The hypotheses presented as a part of research question 1 indicate the expected direction of the effects, which reflects the theoretical framework.

The variables for accessible and inaccessible drinking water have contrasting effects on the level of stunting. There are correlations between piped drinking water and the level of stunting, and unimproved drinking water and the level of stunting. Piped drinking water represents accessible drinking water and affects stunting directly, indicating a lower level of stunting. Conversely, inaccessible drinking water, as represented by unimproved drinking water, indicates an increased level of stunting. The findings show that accessibility has a direct effect on stunting, which is justified by relating it to the theory of energy balance. Although drinking water is not included in the theory of energy balance, the required efforts associated with inaccessible drinking water corresponds to the effects of energy output. In opposition, the lack of effort associated with accessible drinking water makes it possible to use the remaining energy for other purposes.

6.2. The correlation between quality of drinking water and stunting

Research question 2 questions if there is a correlation between quality of drinking water and stunting. Similar to research question 1, it is an initial step to establish whether or not significant effects based on quality exist. The hypotheses connected to research question 2 indicate the expected direction of the effects based on the findings from previous studies and the theoretical framework.

The previous studies indicates that the quality of drinking water has a significant effect on health outcomes (Edokpayi et al., 2018; Hunter et al., 2010; Jain, 2012). However, the hypotheses for research question 2 have been rejected. The results show contrasting effects of undetermined quality drinking water on stunting, indicating inconsistencies within the category of undetermined quality.

The assumption that safely managed drinking water was the most significant variable for explaining the variation in stunting was rejected. The assumption was based on previous studies on the health risks connected to inaccessible and contaminated drinking water. Although safely managed drinking water is an aim in the SDGs as well, the results presented here do not support the direct effect of safely managed drinking water on stunting.

6.3. The inclusion of drinking water into the physical capacity curve

The objective of research question 3 is to initiate a discussion of whether or not drinking water should be included into the physical capacity curve. The argument of including drinking water into the physical capacity curve is strengthened based on the theoretical framework and discussion provided in this thesis. Previous studies and literature support this, linking drinking water to the frequency of diseases (Todaro & Smith, 2012), mortality rate (Haq et al., 2007), and physical abnormalities (Edokpayi et al., 2018), which impacts the ability to do effective work. This initial support of the argument is justified by the rationale that lower nutritional well-being causes lower work capacity by having a higher probability of incapacitating diseases and lower energy input. How access and quality of drinking water affects stunting, as a measure for nutritional well-being, is therefore important.

The significant effect of accessible drinking water is connected to the theory of energy balance, which is the foundation for the physical capacity curve. The benefit of accessible drinking water is the conservation of energy, which in turn can be used for productive work. The physical capacity curve is based on the theory of energy balance, and this logic contributes to the interest of drinking water in connection to nutrition and work capacity. Accessible drinking water can also be regarded as a measure of development. Accessible drinking water implies more advanced infrastructure and higher level of general development. Drinking water can then provide a context to the physical capacity curve rather than supporting an inclusion of drinking water into the theory of the physical capacity curve.

The interaction terms have proven that high quality drinking water and available nutrition have a significant effect on stunting. The effect of undetermined quality of drinking water and available nutrition is less clear. It is reasonable to assume that an increase in the variable limited drinking water causes the variable minimum acceptable diet to affect stunting negatively, indicating a decrease in the level of stunting. Although limited drinking water is also a measure for inaccessible drinking water, the variable ensures high quality drinking water, which is linked to better sanitation (Hunter et al., 2010; Jain, 2012) and less frequency of diseases (Edokpayi et al., 2018; Todaro & Smith, 2012). It is evident that both high quality drinking water and available nutrition affects stunting. Based on these results we can not conclude with a strengthening of the argument of including drinking water into the physical capacity curve. However, in light of the other findings the results are of great interest and would serve a purpose in being studied further.

This thesis has been an initial approach to the question of whether or not drinking water should be included into the physical capacity curve. This study has proven the importance of drinking water in the context of nutrition and nutritional well-being, supporting the inclusion of drinking water into the physical capacity curve. However, based on one study alone we must be cautious not to draw premature conclusions. Modifying an already established theory requires a higher level of certainty, and this can be improved with further empirical studies.

6.4. Final comments

While there exists some literature on the interdependence of nutrition and water, theories within the social sciences generally experience a knowledge gap by omitting water in theories regarding nutrition. There is a shortage of theories within development studies that include this relationship specifically. During the research phase, there were little mentions of water in nutrition-related literature if any at all. Although it is accepted that both nutrition and drinking water are needed for sustaining the human body, how they interact and their place in theories should be expanded upon.

This paper does not aim to provide policy recommendations. However, by using the findings that have been presented it is possible to identify areas that have a potential for improvement. The term improvement used here refers to actions that may lower the probability of stunting in the population. Due to accessibility of drinking water having a high significance for stunting, this is the main area of improvement. Prioritizing the development of better infrastructure and water security for the population would be reasonable measures to ensure universal access to drinking water, and consequently lowering the probability of stunting in the population.

6.5. Further research

An interesting possibility for future research is to measure the effects of proven contaminated drinking water. Using the same research design and methodology but replacing undetermined quality drinking water with low quality drinking water may produce different results. This requires more data, which is challenging to collect for multiple reasons. Data collection on proven contaminated drinking water demands extensive resources as well as there may exist private agendas in disclosing pollution in drinking water sources. However, measuring the effect of contaminated drinking water can provide a clear contrast to high quality drinking water. Furthermore, future studies may benefit from making contrasting hypotheses. While this thesis focuses more on the effects of accessible and high quality drinking water, it could be interesting to focus on the effects of inaccessible and low quality drinking water on stunting.

It would be of interest to do case studies on the effect of drinking water on stunting, on a national level, similar to the study by Choudhary, Schuster, Brewis and Wutich (2021) for India. The advantage of national level case studies is the possibility to easier compare countries and add context for the results. Case studies would also be beneficial to analyse the type of resources different countries or societies have to obtain safely managed drinking water. Resources can range from economic advantages, such as national wealth and investment in public goods, to knowledge about health, nutrition and sanitation. Comparing a few case studies also requires less data, which has proven to be a challenge in the global-level dataset used in this thesis. A possibility to secure representative results could therefore be to compare national level case studies across regions or level of development. Alternatively, further research could benefit from more extensive datasets.

Future research on the topic would also benefit from examining the role of gender. In societies operating with traditional gender roles, women are more often inclined to do household tasks while men do income-generating tasks. Thus, the women would also be responsible for collecting drinking water, given that the drinking water source is not located in immediate proximity. Furthermore, women would have the opportunity to consume more drinking water due to them being by the drinking water source. The question would then arise if women in these societies have a lower probability of giving birth to a stunted child. Using panel data for societies that have or are undergoing a transition from traditional gender roles to a more equal distribution of tasks, where drinking water is not located in immediate proximity, this effect could be examined. These societies would also be of great interest to study to assess any difference in work capacity based on gender and see if there is a gendered physical capacity curve. A relevant measure would in this case also be the distribution of water within the household. Although surveys would more efficiently collect (quantitative) data regarding these relationships, it would be beneficial to combine this method with qualitative assessments. By conducting observations of for example families, one would improve the results by discovering dynamics between individuals within the households.

7. References

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., De Onis, M., . . . Martorell, R. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The lancet, 382*(9890), 427-451.

Choudhary, N., Schuster, R. C., Brewis, A., & Wutich, A. (2021). Household Water Insecurity Affects Child Nutrition Through Alternative Pathways to WASH: Evidence From India. *Food and Nutrition Bulletin, 42*(2), 170-187. doi:10.1177/0379572121998122

Dasgupta, P. (1997). Nutritional status, the capacity for work, and poverty traps. *Journal of econometrics, 77*(1), 5-37.

Dasgupta, P., & Ray, D. (1986). Inequality as a determinant of malnutrition and unemployment: Theory. *The Economic Journal, 96*(384), 1011-1034.

Dasgupta, P., & Ray, D. (1990). Adapting to Undernourishment: The Biological Evidence and its Implications203. In A. S. Jean Drèze (Ed.), *The Political Economy of Hunger: Volume 1: Entitlement and Well-being* (Vol. 1, pp. 191-246).

de França Doria, M. (2010). Factors influencing public perception of drinking water quality. *Water policy, 12*(1), 1-19.

Edokpayi, J. N., Rogawski, E. T., Kahler, D. M., Hill, C. L., Reynolds, C., Nyathi, E., . . . Bessong, P. (2018). Challenges to sustainable safe drinking water: a case study of water quality and use across seasons in rural communities in Limpopo province, South Africa. *Water, 10*(2), 159.

Engell, D. (1988). Interdependency of food and water intake in humans. *Appetite, 10*(2), 133-141.

Freckleton, R. P. (2011). Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. *Behavioral Ecology and Sociobiology, 65*(1), 91-101.

Global Nutrition Report. (2021). *Glossary*. Retrieved from Global Nutrition Report https://globalnutritionreport.org/reports/2021-global-nutrition-report/glossary/

Haq, M., Mustafa, U., & Ahmad, I. (2007). Household's willingness to pay for safe drinking water: a case study of Abbottabad district. *The Pakistan Development Review*, 1137-1153.

Hill, J. O., Wyatt, H. R., & Peters, J. C. (2013). The importance of energy balance. *European endocrinology, 9*(2), 111.

Hunter, P. R., MacDonald, A. M., & Carter, R. C. (2010). Water supply and health. *PLoS medicine, 7*(11), e1000361.

Jain, R. (2012). Providing safe drinking water: a challenge for humanity. In (Vol. 14, pp. 1-4): Springer.

Jha, R., Gaiha, R., & Sharma, A. (2009). Calorie and Micronutrient Deprivation and Poverty Nutrition Traps in Rural India. *World development, 37*(5), 982-991. doi:https://doi.org/10.1016/j.worlddev.2008.09.008

Lang, D. (2015). Opportunities to assess factors contributing to the development of the intestinal microbiota in infants living in developing countries. *Microb Ecol Health Dis, 26*(1), 28316-28316. doi:10.3402/mehd.v26.28316

Leibenstein, H. (1978). *General X-efficiency theory and economic development*: Oxford University Press.

Luh, J., Christenson, E., Toregozhina, A., Holcomb, D., Witsil, T., Hamrick, L., . . . Bartram, J. (2015). Vulnerability assessment for loss of access to drinking water due to extreme weather events. *Climatic Change, 133*. doi:10.1007/s10584-015-1493-0

McKay, N. J., Belous, I. V., & Temple, J. L. (2018). Increasing water intake influences hunger and food preference, but does not reliably suppress energy intake in adults. *Physiology & behavior, 194*, 15-22.

Popkin, B. M., D'Anci, K. E., & Rosenberg, I. H. (2010). Water, hydration, and health. *Nutrition reviews, 68*(8), 439-458.

Ray, D. (1992). *Labor markets, adaptive mechanisms and nutritional status*: Boston University, Institute for Economic Development.

Ray, D. (1998). *Development economics*. Princeton, N.J: Princeton University Press.

Ray, D., & Streufert, P. A. (1993). Dynamic equilibria with unemployment due to undernourishment. *Economic Theory, 3*(1), 61-85.

Sawka, M. N., Cheuvront, S. N., & Carter, R. (2005). Human water needs. *Nutrition reviews, 63*(suppl_1), S30-S39.

The United Nations Childrend’s Fund (UNICEF) (2019). Child Nutrition. Retrieved from https://data.unicef.org/topic/nutrition/child-nutrition/

The United Nations Childrend’s Fund (UNICEF) (a). Height-for-age <-2 SD (stunting), Survey Estimates. Retrieved from https://data.unicef.org/indicator-profile/NT_ANT_HAZ_NE2/

The United Nations Childrend’s Fund (UNICEF) (b). Proportion of population using limited drinking water services. *Indicator profiles.* Retrieved from https://data.unicef.org/indicator-profile/WS_PPL_W-L/

The United Nations Childrend’s Fund (UNICEF) (c). Proportion of population using safely managed drinking water services. *Indicator profiles.* Retrieved from https://data.unicef.org/indicator-profile/WS_PPL_W-SM/

The United Nations Childrend’s Fund (UNICEF) (d). Proportion of population using piped drinking water sources. *Indicator profiles.* Retrieved from

The United Nations Childrend’s Fund (UNICEF), World Health Organization (WHO), & The World Bank (WB) (2019). Levels and trends in child malnutrition: key findings of the 2019 edition of the Joint Child Malnutrition Estimates. *World Health Organization: Geneva, Switzerland*.

The United Nations Childrend’s Fund (UNICEF), World Health Organization (WHO), & The World Bank (WB) (2021). *Technical Notes from the background document for country consultations on the 2021 edition of the UNICEF-WHO-World Bank Joint Malnutrition Estimates*. Retrieved from https://data.unicef.org/resources/jme-2021-country-consultations/

The World Bank (WB) (2018). World Development Indicators: The Story. Retrieved from https://datatopics.worldbank.org/world-development-indicators/stories/world-development-indicators-the-story.html

The World Bank (WB) (2022). Metadata Glossary. Retrieved from https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SH.H2O.SAFE.ZS

The World Bank (WB) (a). *World Development Indicators*. Retrieved from: https://databank.worldbank.org/source/world-development-indicators

The World Bank (WB) (b). World Bank Country and Lendig Groups. Retrieved from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

Todaro, M. P., & Smith, S. C. (2012). *Economic development* (11th Edition ed.). Boston: Eddison-Wesley.

United Nations Environment Programme (UNEP) (2021). *Making Peace with Nature: A scientific blueprint to tackle the climate, biodiversity and pollution emergiencies.* Retrieved from Nairobi: https://www.unep.org/resources/making-peace-nature

https://data.unicef.org/indicator-profile/WS_PPL_W-P/

Vilcins, D., Sly, P. D., & Jagals, P. (2018). Environmental risk factors associated with child stunting: a systematic review of the literature. *Annals of global health, 84*(4), 551.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “p< 0.05”. In (Vol. 73, pp. 1-19): Taylor & Francis.

World Health Organization (WHO). Lenght/height-for-age. Retrieved from https://www.who.int/tools/child-growth-standards/standards/length-height-for-age

World Health Organization (WHO) (2014). *Global nutrition targets 2025: Stunting policy brief*. Retrieved from

World Health Organization (WHO) (2021). *Progress on household drinking water, sanitation and hygiene 2000-2020: five years into the SDGs* (9240030840). Retrieved from

World Health Organization (WHO), & The United Nations Childrend’s Fund (UNICEF) (2018). *JMP Methodology: 2017 update & SDG baselines*. Retrieved from https://washdata.org/sites/default/files/documents/reports/2018-04/JMP-2017-update-methodology.pdf

World Health Organization (WHO), & The United Nations Childrend’s Fund (UNICEF) (2021). *Metadata: SDG global indicator 6.1.1*. Retrieved from https://washdata.org/sites/default/files/2022-01/jmp-2021-metadata-sdg-611.pdf

Appendix 1

```
#the packages used:
library(dplyr)
library(ggplot2)
library(GGally)
library(plm)

#I do not make subsets of the original datasets because I
will not interact with them alone
#Merging the datasets for stunting (Child Nutrition) and
water (WASH). I will call the merged dataset primary.orig

primary.orig <- merge(x = stunting2,
                      y = wash,
                      by.x = c("geogarea", "year",
"country"), #selecting the key variables to merge with from
the child malnutrition dataset
                      by.y = c("geogarea", "year",
"country"), #selcting the same variables from the WASH
dataset
                      all = TRUE) #including all variables
from both datasets
primary.orig

#Merge the datasets for the controlled variables with the
primary.orig dataset into a new dataset: allvar
allvar <- merge(x=primary.orig,
                y=minacceptablediet, #merging with the data
for minimum acceptable diet
                by.x= c("year","country", "geogarea"),
                by.y= c("year","country", "geogarea"),
                all= TRUE)
```

```
allvar <- merge(x=allvar,
                y=gdp, #merging with the data for GDP per
capita
                by.x= c("year","country", "geogarea"),
                by.y= c("year","country", "geogarea"),
                all= TRUE)
allvar <- merge(x=allvar,
                y=education, #merging with the data for
attained education
                by.x= c("year","country", "geogarea"),
                by.y= c("year","country", "geogarea"),
                all= TRUE)
allvar<- merge(x=allvar,
               y=mortality, #merging with the data for
mortality; death rate and life expectancy
               by.x= c("year","country", "geogarea"),
               by.y= c("year","country", "geogarea"),
               all= TRUE)

#Creating a subset of allvar that I can interact with
allvar1<-allvar

#Selecting only the relevant variables
allvar1<-allvar1%>%
  dplyr::select(year=year,
                geogarea=geogarea, #this is the country code.
                country=country, #this is the country name
which will be used for clustering later
                stunting=stunting,
                limitedDW=limitedDW,
                unimprovedDW=unimprovedDW,
                pipedDW=pipedDW,
                safelymanagedDW=safelymanagedDW,
                minaccdiet=minaccdiet,
```

```
                gdp=gdp,
                edu=edu,
                deathr=deathr,
                lifeexp=lifeexp)

#making the descriptive statistics (mean, maximum value,
minimum value)
summary(allvar1)
#running a function for standard deviation since it is not
included in the summary() function
sd(allvar1$stunting, na.rm=TRUE)
sd(allvar1$limitedDW, na.rm=TRUE)
sd(allvar1$unimprovedDW, na.rm=TRUE)
sd(allvar1$pipedDW, na.rm=TRUE)
sd(allvar1$safelymanagedDW, na.rm=TRUE)
sd(allvar1$minaccdiet, na.rm=TRUE)
sd(allvar1$gdp, na.rm=TRUE)
sd(allvar1$edu, na.rm=TRUE)
sd(allvar1$deathr, na.rm=TRUE)
sd(allvar1$lifeexp, na.rm=TRUE)

#Checking for collinearity between independent variables
collin<-allvar1
collin<-collin%>%
  dplyr::select(limitedDW=limitedDW,
                unimprovedDW=unimprovedDW,
                pipedDW=pipedDW,
                safelymanagedDW=safelymanagedDW,
                minaccdiet=minaccdiet,
                gdp=gdp,
                edu=edu,
                deathr=deathr,
                lifeexp=lifeexp)
collinplot<-collin[,1:9]
```

```
ggpairs(collinplot) #creating a visualization that appears in
my plots. Contains both numeric information and scatterplots

#Clustering the data based on country and year into the new
dataset D
D<-pdata.frame(allvar1, index=c("country", "year"),
drop.index=TRUE, row.names=TRUE)
head(attr(D, "index"))

#Running the multiple linear regression for research
questions 1 and 2
model1<-
lm(stunting~safelymanagedDW+pipedDW+limitedDW+unimprovedDW,
data = D)
summary(model1) #getting the data output for the tables
model2<-
lm(stunting~safelymanagedDW+pipedDW+limitedDW+unimprovedDW+mi
naccdiet+edu+deathr+lifeexp, data=D)
summary(model2)
model3<-
lm(stunting~safelymanagedDW+pipedDW+limitedDW+unimprovedDW+gd
p+edu+deathr, data = D)
summary(model3)

#Making the multiple linear regression with the interaction
terms for research question 3
model4<-
lm(stunting~safelymanagedDW*minaccdiet+pipedDW*minaccdiet+lim
itedDW*minaccdiet+unimprovedDW*minaccdiet, data=D)
summary(model4)
model5<-
lm(stunting~safelymanagedDW*minaccdiet+limitedDW*minaccdiet,
data=D)
summary(model5)
```

```
model6<-
lm(stunting~pipedDW*minaccdiet+unimprovedDW*minaccdiet,
data=D)
summary(model6)

#Making the scatterplots with the regression line for
stunting and each type of drinking water
ggplot(D, aes(x=pipedDW, y=stunting)) +
  geom_point() +
  theme_bw() +
  geom_point(position="jitter") +
  geom_smooth(method ='lm', color="red")
ggplot(D, aes(x=safelymanagedDW, y=stunting)) +
  geom_point() +
  theme_bw() +
  geom_point(position="jitter") +
  geom_smooth(method ='lm', color="red")
ggplot(D, aes(x=limitedDW, y=stunting)) +
  geom_point() +
  theme_bw() +
  geom_point(position="jitter") +
  geom_smooth(method ='lm', color="red")
ggplot(D, aes(x=unimprovedDW, y=stunting)) +
  geom_point() +
  theme_bw() +
  geom_point(position="jitter") +
  geom_smooth(method ='lm', color="red")
```

NTNU

Norwegian University of
Science and Technology