Selection and prioritization of factors contributing to overweight and obesity based on country statistics

To what extent can factors that are identified by literature as contributing to overweight and obesity prevalence be confirmed and prioritized based on country statistics?

IN·DEI·NOMINE·FELICITER

Radboud University

Andreas van der Schrier

1004745

Master's Thesis in Economics

Radboud University

January 23rd, 2022

Abstract

The literature provides contradicting information on factors possibly causing overweight and obesity. Therefore, this study investigates factors, as identified in literature, on their significance and relative importance, which are investigated by cross-sectional analyses on data from 189 countries using multiple linear regression.

Worldwide trends in overweight and obesity are confirmed. The study includes complete case analyses and analyses on imputed datasets by predictive mean matching.

The regression is significant for all models for all datasets investigated. Sugar supply and fat supply, the human development index (HDI) and urbanization are all significant in explaining overweight prevalence, with sugar supply and fat supply dominating. Physical inactivity is not associated with overweight prevalence. The KOF Globalization Indices are less successful in explaining overweight prevalence, except for the social dimension. The latter can replace HDI without causing too much change in the regression results of other dominant factors, but HDI performs better.

The results suggest that the increase in a country's development, expressed by a higher HDI and level of urbanization, changes the food environment resulting in a positive energy balance. More energy is consumed than expended, most likely caused by excessive calorie intake rather than a lack of physical activity.

Acknowledgements

The author likes to thank Dr. J. Huisman for her valuable feedback and positive interactions during the process of writing this thesis.

Table of contents

List of figures

Figures in main text:

Figures in Appendix A:

Figures in Appendix C:

Figures in Appendix D:

List of tables

Tables in main text:

Tables in Appendix A:

Tables in Appendix B:

Tables in Appendix C:

Tables in Appendix E:

Glossary

Body Mass Index	-	Body Mass Index (BMI) is a person's weight in kilograms (kg) divided by the height squared (m^2) according to the United States Centers for Disease Control and Prevention (CDC, 2021a).
Obesity	-	Body Mass Index ≥ 30 kg/m^2 according to the definition of the World Health Organization (WHO, 2021a).
Overweight	-	Body Mass Index ≥ 25 kg/m^2 (WHO, 2021a).
Developing countries	-	The developing countries are defined in this study in accordance with page 130 of the World Economic Outlook (IMF, 2018) as "all those that are not classified as advanced economies". The number of developing countries in total is 155. Countries are classified by the International Monetary Fund (IMF) to degree of development based on three criteria: (1) per capita income level, (2) export diversification—so oil exporters that have high per capita GDP would not make the advanced classification because around 70% of its exports are oil, and (3) degree of integration into the global financial system. Developing countries have a per capita income level below the threshold of $2,700 in 2016 as measured by the World Bank's Atlas method. The definition includes however more than this aspect only and is in that sense not strictly defined.
Developed countries	-	Developed countries are the advanced economies according to the World Economic Outlook classification Table B, p. 132 (IMF, 2018).The number of developed countries in total is 39.

Chapter 1 - Introduction

Between 1980 and 2013, the percentage of people with a Body Mass Index (BMI) larger than 25 kg/m^2, classified as overweight, has increased worldwide from 29% to 37% in men, and from 30% to 38% in women. In 2016, 39% of adults aged eighteen years and older were overweight and 13% of adults aged eighteen years and older had a BMI of over 30 kg/m^2 classifying as obese. This adds up to 1.9 billion adults who are overweight, and of these 650 million adults who are obese (WHO, 2021a).

The observed growth in overweight and obesity affects young and old, rich and poor. Between 1980 and 2013 the prevalence of overweight and obesity has increased from about 17% to 24% for people aged 2-19 years old in developed countries and from about 8% to 13% in developing countries (Ng et al., 2014). In 2016, the vast majority of overweight or obese children live in developing countries, where the rate of increase has been more than 30% higher than the rate of increase in developed countries (WHO, 2021a). In practically every region in the world, apart from parts of sub-Saharan Africa and Asia, there are currently more people that are overweight and obese than underweight. Today more deaths occur because of overweight and obesity than underweight. Developing countries are especially hit hard, since they already must deal with multiple problems of their own, for example infectious diseases, undernutrition and childhood stunting. They now must also deal with the burden of an increase in people who are overweight and obese (WHO, 2021a). This combination is often referred to as the 'double burden' in developing countries.

Overweight and obesity clearly have a negative impact on society. They are known causes for diabetes, cardiovascular disease (CVD), cancer and other illnesses like liver disease, infection and depression and result in increasing healthcare costs (WHO, 2021a). The estimated medical costs of obesity in the United States alone accounted for $149.4 billion USD in 2014 (Kim & Basu, 2016). The trend of increasing healthcare expenditure linked to increasing overweight and obesity rates is recognized all over the world (Shekar & Popkin, 2020). But increasing costs associated with overweight and obesity are not only caused by increasing healthcare expenditures. Obesity also results in increasing costs due to lost productivity and foregone economic growth as a result of lost work days, lower productivity at work, mortality and permanent disability (Tremmel, Gerdtham, Nilsson & Saha, 2017).

The World Bank calls for action before it is too late, considering that the rate at which overweight and obesity grows is alarming, and given its severe consequences to society. The World Bank calls overweight and obesity, in their report by Shekar and Popkin (2020), a ticking time bomb ready to explode with huge impact on society in general, and in particular on people living in developing counties.

The primary cause of overweight and obesity is well understood. The normal physiological response to exceeding energy intake over energy expenditure is weight gain. Overweight and obesity result when this energy imbalance persists (Wyatt, Winters & Dubbert, 2006). Proposed interventions to mitigate the trends of increasing overweight and obesity are logically linked to restoring the balance between calories consumed and calories expended, and thus to factors that contribute to restore the balance between calorie intake and calorie expenditure.

The National Health Service of the United Kingdom (NHS), identifies poor diet, lack of physical activity, genetics and medical reasons as factors causing obesity (NHS, 2019). The United States Centers for Disease Control and Prevention (CDC) calls obesity a complex health issue resulting from a combination of causes and individual factors, including behavior and genetics. Factors affecting overweight and obesity as identified by the CDC include physical activity, dietary patterns, medication use, changes in food and physical activity environment, education, skills and food marketing and promotion (CDC, 2021b).

Different factors contributing to overweight and obesity have been identified in literature. The intake of sugar through sugar sweetened beverages, for example, is often referred to as one of the factors explaining the current increase in overweight and obesity (Hu, F., 2013; Du, Tugendhaft, Erzse & Hofman, 2018; Malik & Hu, F., 2019). However, others reported sugar as controversial in explaining overweight and obesity and/or considered sugar even unlikely to be a factor (Stanhope, 2016; Hu, S. et al., 2018). Urbanization was long considered an explaining factor, but more recent data did not confirm this (Tian, Zhao, Li, Wang & Shi, 2014; Costa-Font & Mas, 2016; Popkin, 2019). Mixed statements were reported with regards to physical activity as explaining factor (Hills, Andersen & Byrne, 2011; Ladabaum, Mannalithara, Myer & Singh, 2014). Ross, Flynn and Pete (2016) could not identify a clear pattern or a consensus with regards to the primary factor causing overweight and obesity based on extensive review of literature. Factors covered by their review included diet, physical activity, education, societal environment and genetics. Ross, Flyn and Peter (2016) put therefore question marks on the funding priority for obesity prevention measures.

As demonstrated in the above, factors published in the literature are sometimes contradicting and are not easy to compare in terms of their relative importance, because they are investigated on different populations and often use different methods of analysis. It is not always clear how and on what criteria data have been selected for inclusion in the study, and with regards to choices made to deal with missing data. The latter may affect the research results (Stavseth, Clausen & Røislien, 2018).

The current research aims to investigate the significance and foremost the relative importance of factors that according to literature are most likely linked to the development of overweight and obesity in a systematic way, based on available data on factors at country level. Using country statistics to investigate the response of the country's populations offers a practical way to investigate the factors and their relative importance, where it is considered that it is likely that populations in countries are exposed to the same factors but to different degrees. It is considered that country statistics can be used to investigate the significance and relative importance of these factors by a cross-sectional study over all countries of the world of which sufficient data is available.

The following research question results:

> *To what extent can factors that are identified by literature as contributing to overweight and obesity prevalence be confirmed and prioritized based on country statistics?*

If it appears possible to confirm and prioritize factors that cause overweight and obesity, the study results may be of use to policymakers in prioritizing on factors to mitigate the overweight and obesity problem.

The report is structured as follows. In *Chapter 2* the literature review is presented. The literature review aims to identify factors that can be linked to the growth of overweight and obesity and presents hypotheses with regards to relevance and relative importance of the factors that are identified. *Chapter 3* defines the research methodology and discusses data and the variables to be included in the study. *Chapter 4* presents the analyses and results. *Chapter 5* provides the conclusions. *Chapter 6* presents a discussion on the research findings and recommendations.

Chapter 2 - Literature review

Overweight and obesity are defined as "abnormal or excessive fat accumulation that may impair health" (WHO, 2021a). Whether a person is of healthy weight, is overweight or obese is classified according to its Body Mass Index (BMI), where BMI is defined as the persons mass divided by its height squared. This study uses overweight and obesity as dependent variables and investigates the significance and relative importance of factors, representing the independent variables in this research, that may contribute to overweight and obesity.

The definition of overweight and obesity has been subject to debate over the last few years. Sharma and Campbell-Scherer (2017) proposed to redefine the WHO definition of overweight and obesity by leaving out the word "may", as this would make clear that people identified to have overweight or obesity are actually sick and thus have a real health problem. This would also acknowledge that not all people with a high BMI are sick. The American Medical Association (AMA) had already declared obesity a disease in 2013 (OMA, 2013), but the debate continued afterwards. Covington (2017) summarized the debate as follows: "defining obesity as a disease invalidates the importance of discipline, proper nutrition, and exercise and enables individuals with obesity to escape responsibility. For others, obesity as a disease is a bridge to additional research, coordination of effective treatment, and increased resources for weight loss". The Canadian Adult Obesity Clinical Practice Guidelines defined overweight and obesity as an illness in 2020. They recommended to treat excess fat as a chronic disease, but this would require the assessment of the actual health status (Labine, 2020). More recently, in 2021, the European Commission classified obesity as a non-communicable disease (Burki, 2021).

The classification of overweight and obesity according to BMI has been topic of controversy as well. The percentage of fat mass in the human body should be measured and used as classification, but the BMI is used instead for practical reasons. Although BMI provides an easy measurable alternative for the direct measurement of fat mass, it is not a very accurate method of measurement. In the wording of Müller and Geisler (2017, p.1256): "BMI is a score rather than objectively measured fat mass. Neither is it biologically sound nor does it reflect a suitable phenotype worthwhile to study. It is obvious that BMI can define neither 'excessive fat accumulation' nor functional impairments related to it." They remind that the BMI originally originates from a statistically significant relationship between BMI and mortality. This relationship shows a U-shaped curve if BMI is plotted against mortality with BMI on the x-axis and mortality on the y-axis. The U-shape indicates a high mortality for both, people with a very low BMI as well as for people with a very high BMI. People with a relatively low mortality score in between. BMIs ranging from 18.5 to 25 kg/m^2 were considered to be in a healthy condition. BMI values between 25 and 30 kg/m^2 and BMIs exceeding 30 kg/m^2 were defined as overweight and obese respectively.

Despite whether overweight and obesity are considered a disease or not, and despite limitations in using BMI as a measure to express health, a high BMI is without doubt indicative of a high risk to health. BMI is a well-accepted way of defining overweight and obesity and is frequently used as a measure to investigate the impact of overweight and obesity on society (Tremmel, Gerdtham, Nilsson & Saha, 2017; WHO, 2021a; Shekar & Popkin, 2021).

In this study, overweight and obesity prevalence is taken as dependent variable, because it is included in almost all country statistics and because it provides the possibility to couple overweight and obesity trends with a variety of other variables that may be of interest to the current study.

Overweight and obesity result from a sustained positive energy balance over a longer period of time, which is caused by an excess of calorie intake compared to calorie expenditure (Romieu et al., 2017). But many factors can potentially affect the energy balance (WHO, 2021a; NHS, 2019; World Bank, 2020; CDC, 2021). Factors that affect the energy balance negatively, causing weight gain and potentially overweight and obesity, include eating food rich in sugar and fat, drinking sugar sweetened beverages and alcohol, lacking physical activity, behavior and genetics, medication use, stunting, fast food marketing and promotion, globalization, urbanization and economic development. Factors that may affect the energy balance in a positive way include education, policies and interventions to promote a healthier diet and an increase in physical activity.

In the below, findings from the different literature on the impact and relative importance of factors affecting overweight and obesity are presented and discussed. Hypotheses are developed to provide answers on the question whether factors that may cause overweight and obesity can be identified and prioritized based on statistics at country level.

Smethers and Rolls (2018) note that the question whether energy intake would be excessive and cause weight gain, or not, depends not only on the quantity of food consumed but also on the energy density of the components in the food that is consumed. The energy density of food may well be a factor in explaining overweight and obesity. Energy density is defined as the number of kilocalories per gram, and differs by type of food. From all macronutrients, which refers to fats, carbohydrates and proteins, fat has the greatest energy density of about 9 kcal/g. Carbohydrate and proteins are less energy dense with an energy density of around 4 kcal/g. The energy density of the food depends also on its water content. Water content contributes in weight and volume, but does not add any energy, as the energy density of water is 0 kcal/g (Rolls, 2017). Alcohol in pure form, on the other hand, is a fluid with a relatively high energy density (7 kcal/g). The National Health Service of the United Kingdom (2019) discourages eating large amounts of ultra-processed or fast food, that are high in sugar and fat, and drinking too much sugar sweetened beverages and alcohol, because they provide too many calories per gram consumed. That the current increase of overweight and obesity prevalence may be explained by a higher energy density of the food, is supported by the fact that food supply data show an increase of the energy available for consumption in all countries (WHO, 2003). More food is currently available for consumption than ever before. A clear association between the increase of food supply and the increase of overweight and obesity in western countries exists for adults between 39 and 64 years old when food supply data are compared with BMI recordings (Silventoinen et al., 2004). This suggests that nutrients with a high energy density would be good candidates for inclusion in the current study. However, despite its relatively high energy density, the role of sugar in explaining overweight and obesity is still controversial. Sugar may have a direct effect on the epidemics of metabolic disease, including type 2 diabetes, but its role in causing overweight and obesity prevalence is less clear, mainly in view of still missing clinical diet intervention studies which would be required to provide the final evidence (Stanhope, 2016). Other authors, like Hu, F. (2013), Du, Tugendhaft, Erzse and Hofman (2018) and Malik and Hu, F. (2019) emphasize the importance of sugar sweetened beverages in explaining the current increase in overweight and obesity prevalence. Over time, public policies and regulatory strategies have undoubtedly been introduced, such as sugar tax, despite a strong opposition from the beverage industry. Today, even the Coca-Cola Company, one of the leading carbonated soft drink companies of the world (Reiff, 2020), publishes on its own website that they actively work to reduce sugar in their products (The Coca-Cola Company, 2021). Clinical studies to investigate the relative contribution of carbohydrates, fats and protein in human bodies are required over longer periods of time to provide further clarity, but such studies are not easy to conduct for ethical reasons. Hu, S. et al. (2018) therefore investigated the relative contribution of carbohydrates, fats and protein on body fat accumulation in mice. They found that an excessive intake of fats increased body fat content, rather than carbohydrates or protein. Although the findings were found in research on mice, it is noted that mice are known to have many similarities to humans in their physiology and metabolism: “the evidence it provided (the study on mice) is a good clue to what the effects of different diets are likely to be in humans” as explained by Speakman (Cohut, 2018). Hu, S. et al. (2018) further suggest that fats are playing the major role in weight gain, mainly because they appeal to the brain’s reward system, stimulating to consume excessive amount of calories. But, not all fats are unhealthy. Three different types of fats are distinguished: unsaturated-, saturated-, and trans fats. The unsaturated fats are the good fats, as they can improve cholesterol levels, stabilize heart rhythms and ease inflammation. Saturated fats and specifically trans fats are bad types of fat. Saturated- and trans fats are bad because they increase the amount of ‘bad’ cholesterol (low-density-lipoprotein), while at the same time lowering the amount of the ‘good’ cholesterol (high-density-lipoprotein). Eating too much saturated- and trans fats can cause among others coronary disease and diabetes (Mensink, Zock, Kester & Katan, 2003; Mozaffarian et al., 2004; Harvard School of Public Health, n.d.). Fast food, like burgers, fries and fried chicken, contain a relatively large amount of saturated fat and are relatively unhealthy. Literature has indeed linked higher rates of fast food consumption to higher BMIs and increasing rates of obesity (Jeffery, Baxter, McGuire & Linde, 2006; Frazer, Clarke, Cade & Edwards, 2012; Garcia, Sunil & Hinojosa, 2012). A contributing factor is that fat has a low satiety property (Golay & Bobbioni, 1997). In conclusion, (saturated) fats in food can easily result in overconsumption and weight gain, because they appeal to the brain’s reward system, because they have a low satiety property and because they have a high energy density.

In view of the above, it is considered that food supply, and in particular sugar and fat supply, should be included in the current study to investigate their significance and relative importance as factors explaining differences in overweight and obesity prevalence. It is considered better to focus on the components fat and sugar supply separately, rather than on total food supply as a container variable including among others fat and sugar supply. Separation offers the possibility to differentiate between fat and sugar with regards to their relative contribution to overweight and obesity prevalence. Moreover, fast and processed food are high in sugar and fat content. It is decided to include sugar and fat supply as independent variable in current study. It is hypothesized that sugar and fat supply are positively associated with overweight prevalence (Hypothesis 1A). Fat supply is considered dominant over sugar supply (Hypothesis 1B) because of its high energy density, its low satiety property and its appealing effect on the brain's reward system.

Different authors emphasize the importance of economic and technological development in explaining the observed increase in overweight and obesity prevalence in countries. Economic and technological change, together with an increased purchasing power, result in a change of the food environment (Zobel, Hansen, Rossing & Scholten, 2016). Technological developments in food production, processing and distribution means provided global access to a wider variety of foods. People living in particularly urbanized areas started to consume food produced by others rather than by themselves. Large supermarket chains replaced fresh markets as the major source of food. Their market volumes reached values large enough to transform dietary patterns. Marketing and promotion of fast food consumption increased, with most of the efforts concentrated in upcoming developing countries, targeting new markets following the increase in income. Various authors argue that food pricing plays an important role (Drewnowski, 2004; Wright & Aronne, 2012; Dhakal & Khadka, 2021). Energy dense food, high in sugar and fat, provides the lowest cost and highest palatability option to the consumer. They are offered in relatively large portions. Low income households spend a relatively large part of their food budget at convenience stores and fast food restaurants, selling food of lower diet quality. Low income households are found to be the most likely to be obese. Different authors relate the exposure to and the relative ease to access fast food restaurants with an increase of bad food consumption (Davis & Carpenter, 2009; Cooksey-Stowers, Schwartz & Brownell, 2014; Ghosh-Dastidar et al., 2014; Burgoine, Sarkar, Webster & Monsivais, 2018). So-called food deserts develop in low income urban areas, leaving ultra-processed foods and sugar sweetened beverage as the only available food option. Apart from food deserts, food swamps also develop in low income urban areas. Food swamps are characterized by a relatively high number of places selling unhealthy foods, relative to healthier food options. In rural areas, on the contrary, people would still consume mainly home grown foods. People living in urban areas, would also adopt to a more sedentary life style, as their occupations would require less physical activity than occupations in the rural areas. Their physical activity would reduce due to technological change, including the introduction of television, internet and modern means of transport. A number of authors found associations between urbanization and overweight and obesity developments, which they explained by a changing diet pattern, from healthy to ultra-processed food, and a more sedentary life style (Popkin, 2019; Kuddus, Tynan & McBryde, 2020). Costa-Font and Mas

(2016), however, did not find a clear association between overweight and obesity, in their study of the impact of globalization on overweight and obesity with urbanization included as factor. It is decided to include urbanization as a potential factor in explaining differences in overweight and obesity prevalence. It is considered that urbanization has a negative effect on the food environment, resulting in the more consumption of foods relatively high in fat and sugar. It is hypothesized that urbanization is positively associated with overweight prevalence (Hypothesis 2).

Sayon-Orea, Martinez-Gonzalez and Bes-Rastrollo (2011) found contradicting results from their systematic review of 31 papers on the relation between alcohol consumption and body weight. Traversy and Chaput (2015) also called the experimental evidence for an association between alcohol and obesity “mixed”. The National Health Service of the United Kingdom (2019), however, discourages the consumption of too much alcohol as part of the promotion of a healthier lifestyle aiming to prevent overweight and obesity.

Based on the above, it is decided to include alcohol consumption as a potential factor in explaining differences in overweight and obesity prevalence. It is considered that alcohol contributes to overweight and obesity prevalence, in view of its high energy density and assumed association with an unhealthy lifestyle. However, the positive association between alcohol consumption and overweight and obesity prevalence, is likely present under specific conditions only. For example in case of heavy drinking. It would also depend on the type of alcohol that is consumed, where it is noted that beer contains more calories than wine. It is considered unlikely that a hypothesis on the relation between alcohol consumption and overweight and obesity prevalence can effectively be tested at country level, in view of the fact that alcohol consumption is prohibited to a large part of the world’s population for religious reasons. But also because, on the average, excessive alcohol consumption would level out. The relation between alcohol and overweight and obesity prevalence can better be investigated on an individual level. For this reason, alcohol is not included as factor in current analysis.

Different articles link stunting at low age with a larger probability of becoming overweight and obese at later age, caused by a shift in their metabolic properties resulting in storage of fat rather than using fat as the primary source of energy (van der Merwe & Pepper, 2006; Muhammad, 2018). But van der Merwe and Pepper (2006) reported conflicting evidence for the association between stunting and overweight and obesity in South African children, while supporting evidence was found in India. It is concluded that an association may exist, but checking this hypothesis requires study of developments over time, considering the large time period that exists between time of cause and time of effect. Current study follows a cross-sectional approach comparing countries using 2016 as reference year. Stunting is considered outside the scope of the current study.

Hills, Andersen and Byrne (2011) reviewed the relation between physical activity and overweight and obesity in children based on a review of international literature. They reported that a strong relationship exists between physical activity and obesity in children and adolescents. But, they recommended that promotion of a healthy diet and more physical activity would be equally important in overweight and obesity prevention. Ladabaum, Mannalithara, Myer and Singh (2014) examined trends in obesity, abdominal obesity, physical activity, and caloric intake in US adults from 1988 to 2010 and found a significant impact of physical activity level on overweight and obesity in cases of an absence of significant population-level changes in average daily caloric intake.

It is considered that both, diet and physical activity, must be of importance, if accounting for the primary mechanism causing weight gain and thus potentially overweight and obesity, namely the distortion of a healthy energy balance over a longer period of time. Physical activity and diet act on opposite sides of the energy balance. It is decided to include physical activity in the study. It is the only factor that is directly linked to calorie expenditure. It is hypothesized that physical inactivity is positively associated with overweight prevalence (Hypothesis 3A). It is also considered that it is relatively easy to consume more calories than expending them as it requires discipline and effort to compensate for the calories consumed by an increase of physical activity. Therefore overweight is more likely caused by excess calorie intake rather than increased physical inactivity (Hypothesis 3B).

Many papers discuss the impact of global economic development, although considered from somewhat different perspectives (Tian, Zhao, Li, Wang and Shi, 2014; Costa-Font and Mas, 2016; Popkin, 2019). It is considered that globalization, economic development, modernization, and urbanization are interlinked, where it is noted that globalization is defined as the development of an increasingly integrated global economy marked especially by free trade, free flow of capital, and the tapping of cheaper foreign labor markets (Merriam-Webster Dictionary, 2022). It is considered that globalization can affect societies in different ways and can be considered from different perspectives. The KOF Index of Globalization (KOF index) offers a way to look at the developments of the degree of globalization in an economic, social and political dimension (KOF, 2022). A robust association between globalization, obesity and caloric intake, was found in a panel study including 26 countries over the period 1989-2005 (Costa-Font and Mas, 2016). In their study the social globalization dimension provided a robust explanation for the increase of the obese and overweight population and the rise of calorie consumption "alongside alternative explanations of the considered impact of globalization, such as the independent effect of increasing female labor market participation, income inequality and national income, alongside urbanization" (Costa-Font & Mas, 2016, p.130). The relatively low impact of urbanization may be explained by the fact that many studies with regards to the relation between urbanization and overweight and obesity are carried out on too specific data samples, and, for example, include only developed or developing countries, or focus on specific regions or countries only (Popkin, 2019). A study conducted by the NCD Risk Factor Collaboration, which investigated global BMI trends based on the latest data while including almost all countries of the world, found that overweight and obesity are already greater in rural than in urban areas, with the rate of growth in rural areas exceeding the rate of growth in urban areas in specifically the group developing countries. Popkin (2019) clarified that the arguments that were reserved to urban areas in the past, would today also apply to rural areas. Fast food has become available in rural areas, just as the access to cheap mechanized devices for farming and transport. Tian, Zhao, Li, Wang and Shi (2014) had found similar trends in China, based on their logistic regression results on a sample of over 6000 adult residents of 4 provinces of China. From this all, it is considered that globalization does matter, as it results in increased economic, social and political interactions. Interactions that will affect people's food environments and lifestyles. It is hypothesized that globalization through the KOF globalization indices is positively associated with overweight prevalence (Hypothesis 4).

The Human Development Index (HDI) may offer a strong alternative variable for the KOF globalization index. It also includes multiple dimensions (life expectancy, education and standard of living). The health dimension is measured as life expectancy at birth. The education dimension is measured as the mean years of schooling of adults aged 25 and older and the expected years of schooling of children entering school. The last dimension is the standard of living dimension, which is measures by using the logarithm of the gross national income (GNI), to reflect the diminishing importance of income (UNDP, 2022). Ataey, Jafarvand, Adham and Moradi-Asl (2020) investigated the relation between overweight and obesity with the HDI. They reported a significant and positive relationship between the HDI and overweight and obesity prevalence in both females and males for countries in the Eastern Mediterranean Region. Working with HDI as measure has also the advantage that it integrates the economic dimension with the educational dimension, as an interaction may exist between education and economy. This because an interaction was reported between GDP and education in literature (Strand, Vollset & Skirbekk, 2015).

HDI is included in the analysis as it may provide a powerful alternative for the globalization index or its dimensions, as it addresses the peoples economic well-being, their health condition, as well as their state of education at country level which most likely goes together with globalization (international trade, tourism, modernization, etc.). It is hypothesized that HDI is positively associated with overweight prevalence (Hypothesis 5A). Furthermore, it is hypothesized that the Human development index and the KOF globalization indices are well correlated (Hypothesis 5B).

For purpose of completeness, it is noted that it was originally considered to also include the number of fast food restaurants per million inhabitants as a measure for economic interaction and globalization. The factor was calculated as the sum of the number of restaurant from the fast food chains McDonalds, Subway and KFC in countries, where the original number of restaurants were obtained from the companies websites. However, the KOF social dimension offers a better variable as it already includes the number of McDonalds restaurants alongside other social interaction factors (refer to Appendix D). It is decided to not include the number of fast food restaurants per million inhabitants in this study. It is also noted that GDP is excluded as factor because the economic dimension is well-covered already by the economic dimensions in both the HDI and the KOF index.

It is considered that it must be possible to associate at least a number of factors with overweight prevalence and to connect conclusions to these associations with regards to their relative importance. This formally implies that the relationship obtained with multiple linear regression, including all factors, must be significant in explaining overweight prevalence (hypothesis 6). Confirmation of this hypothesis implies a positive answer on the research question, provided that more than one of the factors proves to be significant.

From the literature review the following hypotheses result (*Table 2-1*).

Table 2-1: Hypotheses developed in the framework of this research

No:	Hypotheses on factors identified in literature [1)]	Page[3)]
1A	Sugar and fat supply are positively associated with overweight prevalence[2)]	7
1B	Fat supply dominates sugar supply in explaining changes in overweight prevalence	7
2	Urbanization is positively associated with overweight prevalence	8
3A	Physical inactivity is positively associated with overweight prevalence.	9
3B	Overweight is more likely caused by excess calorie intake rather than increased physical inactivity.	9
4	Globalization through the KOF globalization indices is positively associated with overweight prevalence	10
5A	Human development index is positively associated with overweight prevalence	10
5B	Human development index and KOF globalization indices are well correlated.	10
6	The relationship obtained by multiple linear regression is significant	11

Notes:
1) The validity of the hypotheses will be investigated for their association with overweight and obesity prevalence by multiple linear regression analysis.
2) A strong association exists between overweight and obesity prevalence. The research takes overweight prevalence as the dependent variable (*Section 3.3*).
3) Backgrounds on the selected hypotheses are given on the page indicated.

Chapter 3 - Research methodology

3.1 Introduction

This section describes the methodology adopted to answer the research question:

> *To what extent can factors that are identified by literature as contributing to overweight and obesity prevalence be confirmed and prioritized based on country statistics?*

To answer this question a cross-sectional multiple linear regression analysis is used. The modelling approach, includes the model, as well as the statistical tests required to investigate the validity of the approach are discussed in *Section 3.2*. A complete case analysis is conducted, which means that all countries that miss data in one or more of the factors (independent variables) are excluded from the analysis. The possible impact of leaving data out, is discussed in *Section 3.2*, and further investigated in *Chapter 4*. The statistical tests required to confirm the validity of applying the regression approach are discussed in *Section 3.2*, together with the dominance analysis that is used to assess the relative importance of factors included in the regression analyses. Restrictions with regards to the maximum number of factors (independent variables) than can be included in the analysis and the selection of these factors are discussed in *Section 3.4*. The statistical tests and the dominance analysis are carried out as part of the analyses (*Chapter 4*).

The overweight and obesity trends observed in the world, as already discussed in *Chapter 1*, are verified first. This to confirm that these trends can be reproduced with the own datasets, but also to check for anomalies in the datasets, refer to *Section 3.3*. Overweight prevalence will eventually be selected as the only dependent variable in this study. The choice for overweight prevalence as only dependent variable is also explained in *Section 3.3*. Some analyses are already included in *Chapter 3* because these are needed to decide on the number of independent variables that can be included. The number of independent variables should not be too large, depending on the sample size (number of countries), to still produce reliable regression results (*Section 3.4*).

Chapter 3 stops where the final analyses begin, which means that the final regression analyses and the results are presented in *Chapter 4*.

3.2 Modelling

The multiple linear regression model is applied with (n) independent variables:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + u \quad [1]$$

Where:

Y = dependent variable (overweight prevalence);
β_0 = regression intercept;
$\beta_1 \ldots \beta_n$ = coefficients measuring the change in Y with respect to respectively X_1 to X_n;

$X_1 \dots X_n$ = independent variables X_1 to X_n;
u = error or disturbance term.

Specific characteristics of the independent variables X_1 to X_n included in the model, such as their formal definition, method of measurements and descriptive statistics are given in *Section 3.4*.

The regression model provides the amount of variance in overweight prevalence that can be explained by the independent variables included in the analysis. The relationship obtained by the multiple linear regression will be tested by the so-called F-test. Investigation of the relative importance between variables included is only meaningful if the variables in de model are significant in contributing to overweight prevalence. The relative importance of these variables will be determined by dominance analysis (Azen & Budescu, 2006). In a dominance analysis the additional contribution of a given independent variable is measured by the increase in R^2 that results from adding that independent variable to the regression model. The procedure calculates the relative contribution to the total R^2 for all possible ($2^p - 1$) subsets. For example if the independent variables A, B and C are included, and the relative contribution of variable A to the total variance is to be investigated, a total of 7 subsets is analyzed (A, B, C, AB, AC, BC and ABC).

To apply the multiple linear regression model, different requirements must be fulfilled in order to get the best linear unbiased estimators (BLUE):

- The model must be linear in parameters, which is checked by visual inspection of correlation plots between the variables;
- The sample must present a random sample, for most samples derived from statistical databases this will be the case;
- No perfect collinearity. This condition is checked by inspection of the pairwise correlation between the independent variables, with focus on pairwise correlation and by looking at the Variance Inflation Factor (VIF) with focus on multicollinearity;
- Zero conditional mean. Zero conditional mean is checked visually, by looking at the distribution of error terms in a plot showing the predicted (or fitted) values against the residuals. The average distance of all points to the regression line through the points should be approximately the same. It is noted that heteroskedasticity does not necessarily excludes the possibility of complying with the zero conditional mean condition;
- Heteroskedasticity refers to the variance of the residuals. The variance should be approximately constant for any value of the independent variable. In other words the distance to the regression should not systematically change. This is checked again in the plot showing the predicted (fitted) values against their residuals and by the Breusch-Pagan/Cook-Weisberg test.

The maximum number of independent variables that can be included is restricted by the maximum number of observations available. The regression result may become biased, if too many independent variables are used with having too few observations available (sample size). The recommended ratio between the sample size (number of countries) and the number of independent variables to be included (the factors) is 15 (Schmidt, 1971; Green, 1991; Schneider, Hommel & Blettner, 2010; Austin & Steyerberg, 2015).

The maximum number of countries included in the complete case analysis is 128 (developed and developing countries). The complete case analysis includes all countries that have data available for each factor included. From the 189 countries available, 61 countries are excluded because they miss data in one or more factors. The complete case analysis with developing countries only includes 96 countries, meaning that 93 countries are omitted from the 189 available because they miss data or are not a developing country. Based on the above, the number of factors (independent variables) that can be included are respectively 8 and 6 for the case in which all countries are included and the case with the developing countries only. The number of variables that eventually is included in the regression analysis, and the selection of variables for inclusion are further discussed in *Section 3.4*.

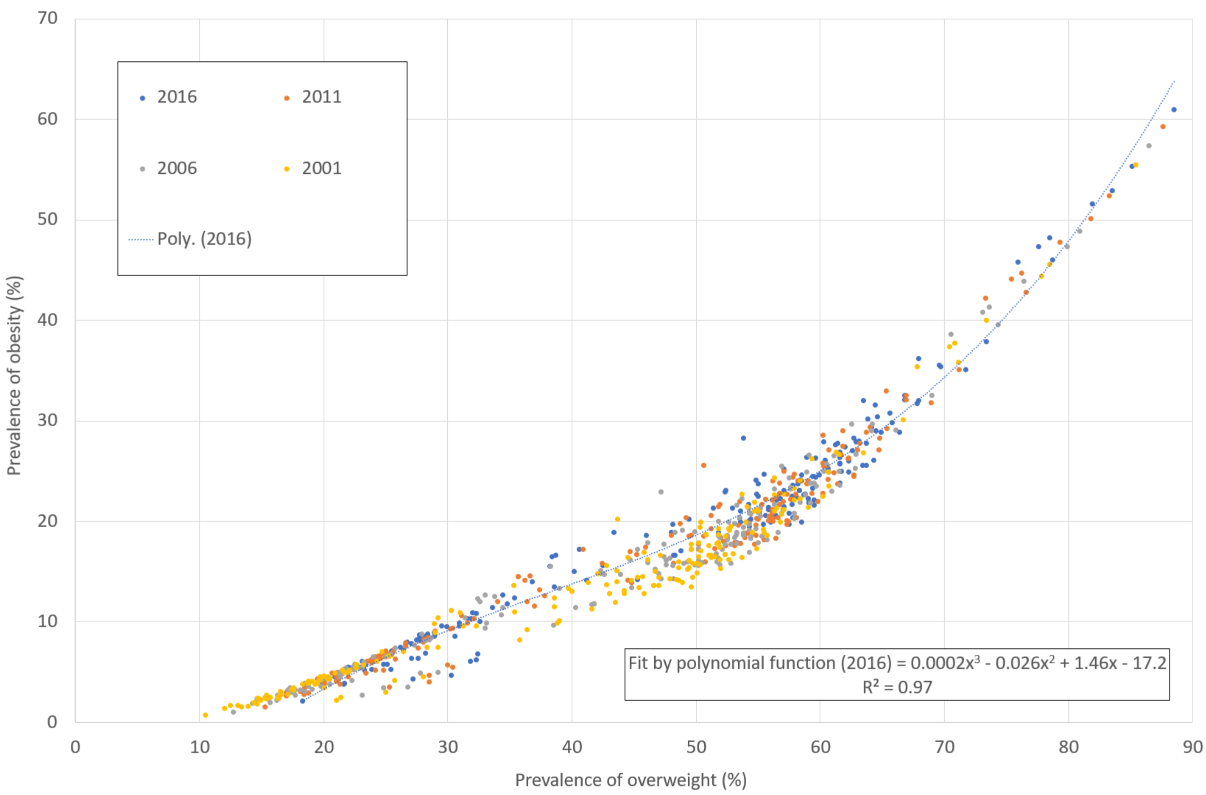
The impact of leaving out countries from the analysis because of missing data will impact the result, with its severeness depending on the reason why the data are missing. If the data are missing completely at random, the data are underpowered meaning that the data are less good in predicting a real effect, but the data remain unbiased. Unfortunately data are seldom missing completely at random, and the missing data may therefore result in biased or unreliable parameter estimates (Lodder, 2014).

Estimated values can be put at positions of missing data, as if these estimated values are real values, to complete the dataset. This is called single imputation, but entering missing data, does not reflect the uncertainty that comes together with the estimation of these missing data. Uncertainty associated with predicting missing data can only be included by doing the imputation process multiple times. In this way each missing value is replaced by a simulated set of possible values (McCleary, 2002), so that many imputed datasets result. This process is called multiple imputation and will be adopted in this research. As a first step in the multiple imputation process, many complete sets of data are generated by imputing the missing values. In the second step a multiple regression analysis is carried out for each of the (many) datasets generated. In the third and last step the results of all the regressions are pooled and analyzed. Multiple methods of imputation are checked with the own dataset (Appendix C).

3.3 Global trends in overweight and obesity prevalence

Overweight and obesity prevalence may both be selected as dependent variables in the regression analysis. However, both variables are based on the same measure (BMI), and are also strongly correlated, with a coefficient of determination (R^2) close to 1 (*Figure 3-1*). Obese people are a subset of the set overweight people, with the only difference being that their BMI number is higher (≥30 instead of ≥25 respectively). If people are not overweight, they cannot be obese. However, countries with a large percentage of overweight people, have also a relatively large percentage of obese people, but the amount of obese people relative to the amount of overweight people increases with the amount of people that are overweight. For example: if about 30% of the people are overweight, about one third of this 30% is obese, if 70% of the people are overweight, about half of this 70% is obese. In view of the above, overweight is selected as the only dependent variable in the regression analysis in *Chapter 4*. For the simple reason that prevention is to be preferred over treatment, if the amount of overweight people remains small, the amount of obese people relative to the people with overweight is small too.

Figure 3-1: Overweight vs obesity prevalence

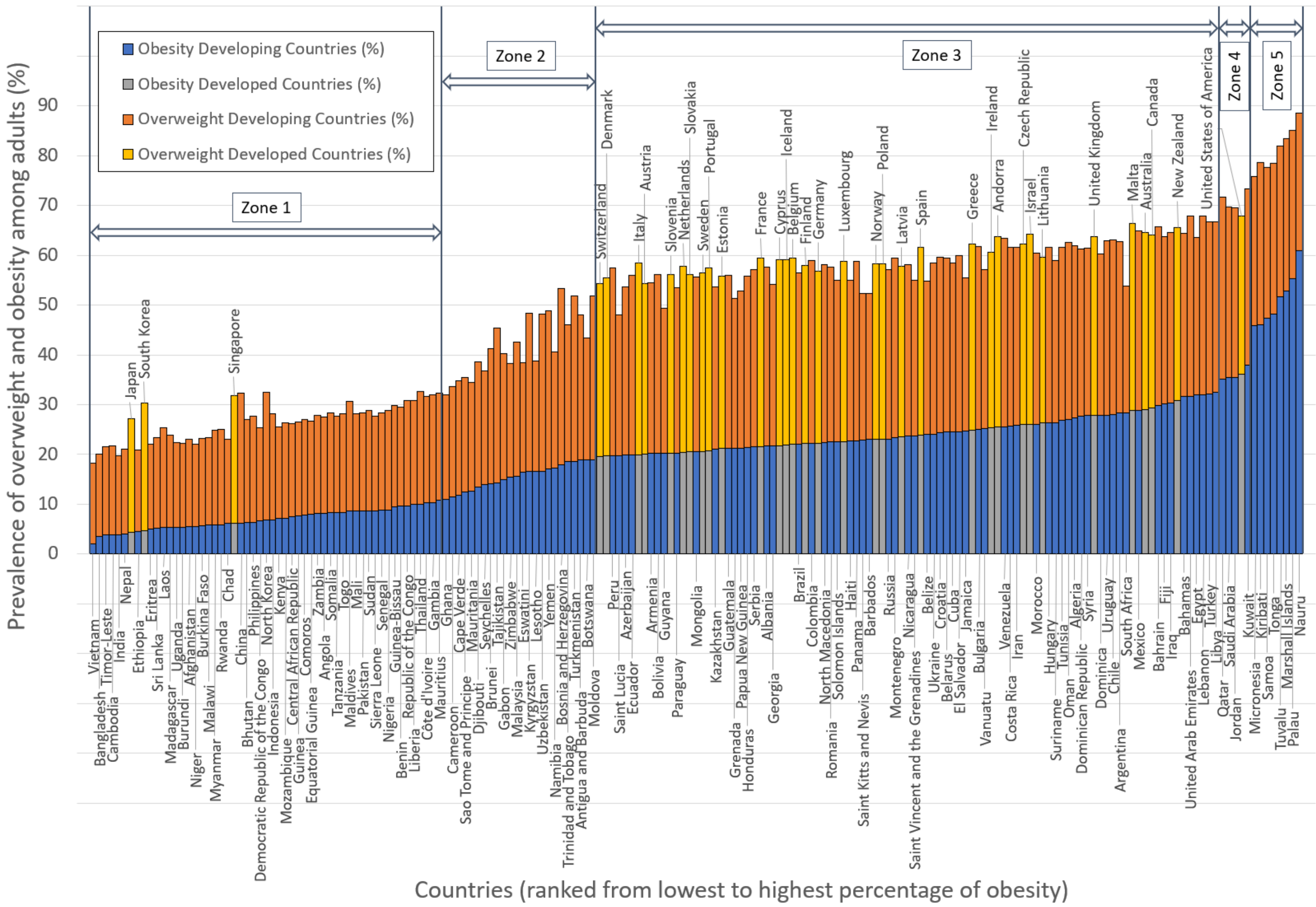


Overweight and obesity prevalence are plotted in increasing order for the year 2016 for the full dataset comprising 189 countries of the world (*Figure 3-2*). The bars in *Figure 3-2* are not cumulative, meaning that for the country Nauru, with largest overweight and obesity numbers, the prevalence of overweight = 88% and the prevalence of obesity = 61%.

A large amount of countries, the first 55 in ranking, are still relatively low in overweight and obesity prevalence (zone 1). Another relatively large group of countries (zone 3) counts 97 countries in total, which have an overweight and obesity prevalence almost double of that in the countries of zone 1. A small group of countries, 24 in total, shows intermediate percentages of overweight and obesity prevalence (zone 2), whereas the countries in zone 4 (5 countries) and in zone 5 (8 countries), show already large to extreme overweight and obesity prevalence.

Countries in the Near and Middle East plot relatively high in zone 3, and plot almost exclusively in zone 4, indicating a relatively high obesity rating (Bahrain, Iraq, United Arab Emirates, Egypt, Lebanon, Turkey and Libya, followed by Qatar, Saudi Arabia, Jordan & Kuwait).

Figure 3-2: Overweight and obesity prevalance for 2016 plotted to increasing obesity



Countries in zone 5 are all Pacific Island States and show extreme overweight and obesity prevalence. WHO (2003) already noted the exceptional position of the Pacific Island States on the obesity ranking in 2003. They report a change in the traditional island diet from fish, root vegetables and coconut to a diet relatively high in saturated fat. According to the WHO report (WHO, 2003, p.44): “Most Pacific countries are food-dependent. A part of good governance for any population is the freedom to choose from a range of nutritious foods. It appears that many people do not get that choice”. Hawley and McGarve (2015) summarize several factors that could be linked to specifically the Pacific countries. These include political and economic change, modernization and migration and urbanization. These factors would have shifted the economy from a more livelihood based economy to a market based economy. Traditional foods such as fish, taro, yams, and indigenous fruits and vegetables are replaced by imported rice, sugar, canned foods, soft drinks, and calorie-dense snack foods. In Samoa and Guam, as much as 80% to 100% of their food is now imported. The foods are not only easy to obtain but also low in cost. The traditionally high-status local foods are now the more expensive ones and therefore saved for special occasions.

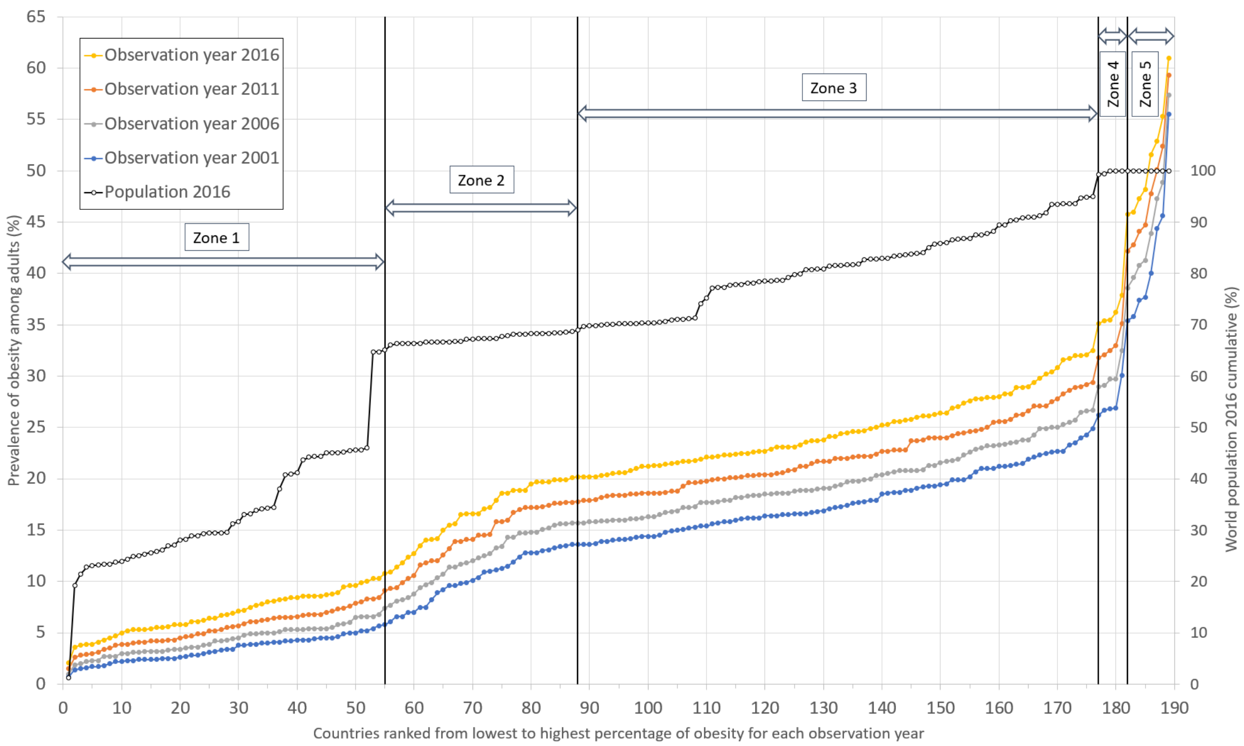
Some Pacific Islands tried to prevent import of low quality food, but stopped these efforts to join the World Trade Organization. In the wording of Swanbrow (2012) on the website of the

University of Michigan: “some Pacific Island nations have attempted to ban the importation of turkey tails and other high-fat food products, but the bans have been lifted in order for these nations to join the World Trade Organization”.

From *Figure 3-2*, it can be observed that developed countries score relatively high in ranking and are mostly found in zone 3, except for Japan, South Korea and Singapore. Japan scores extremely low among the developed countries. Senauer and Gemma (2006) suggested that the lower prevalence of obesity in Japan could be explained by a lower calorie intake, partly caused by relatively high food prices, healthier traditional dietary habits, relative low consumption of animal fat and meat and by a relatively high physical activity. Tsugane (2021) considers that the low mortality rates from heart disease and cancer reflect the low obesity prevalence in Japan and links the low obesity prevalence with a diet that can be characterized by a low red meat intake, a low intake of saturated fatty acids, and a high intake of fish, plant food and non-sugar sweetened beverages like green tea.

To investigate trends in ranking over time, the ranking is plotted for the years 2001, 2006, 2011 and 2016, together with the world population (World Bank, 2021c) plotted cumulatively based on the ranking of 2016, refer to *Figure 3-3*.

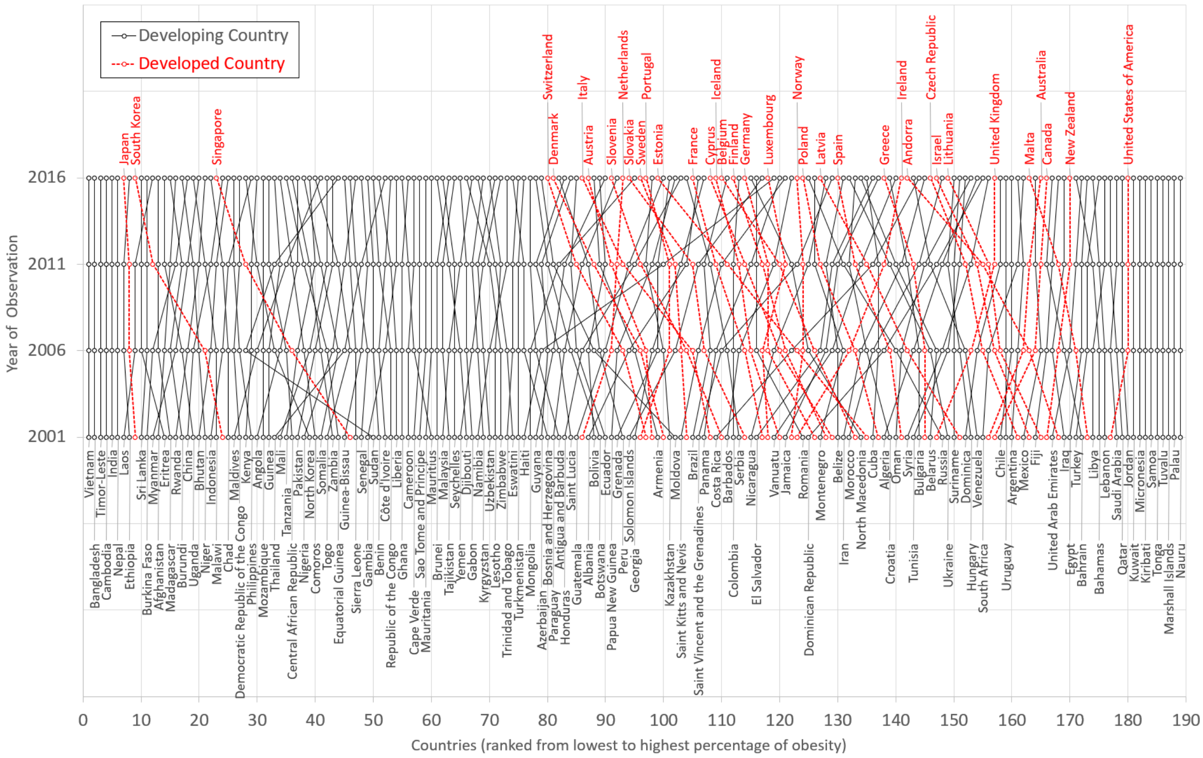
Figure 3-3: Ranking of obesity prevalance among adults over time



It is noted that zones 1, 2, 3, 4 and 5 include respectively 65.1%, 3.9%, 30.4%, 0.6% and <0.01% of the total world population. The majority of the world’s population is located in zone 1 and zone 3.

It is noted that the some countries changed in ranking over the years, but differences are relatively small, refer to *Figure 3-4*.

Figure 3-4: Change in obesity ranking between countries over the years



From *Figure 3-3*, the growth in obesity prevalence occurs in all countries, and at an almost equal rate. Changes in ranking occur over the years, but not drastically (*Figure 3-4*). The sum of the rank numbers of a selection of countries can be considered as a measure to investigate the average change of the relative position of this selection on the rank list. For example if the selection would consist of the first four countries of the list, the sum of rank numbers would be 10 (1+2+3+4). If a few years later, these countries would take the last four positions of the list, the sum of rank numbers would be 750 (186+187+188+189). In the same way the sum of rank numbers of the developing countries in 2001 is compared to the sum of the rank numbers in 2016. The sum increased with 363, of which about 40% occurred in period 2011-2016. This indicates that the increase in obesity prevalence occurs faster and at increasing rate in developing countries in comparison to the developed countries. The observation that the vast majority of overweight children live in developing countries, and that the rate of increase was 30% higher than the rate of increase in developed countries in 2016 (WHO, 2021a) shows that this trend is real and will probably change for the worse. The developing countries will take the lead if the current trend continues.

Despite differences in rate of growth, it should be noted that in all countries obesity numbers are increasing (*Figure 3-3*), but at different rates of growth. The rate of growth increased in 84% of the countries if the period 2011-2016 is compared to period 2006-2011. The remaining 16% present countries (*Table 3-1*) with still increasing obesity numbers, but at smaller rates than before.

Table 3-1: Countries with decreased rate of growth in obesity

Countries with a decreased rate of growth over period 2011-2016 compared to period 2006-2011	Developed countries	Developing countries
Andorra	Andorra	
Belarus		Belarus
Botswana		Botswana
Canada	Canada	
Costa Rica		Costa Rica
Dominican Republic		Dominican Republic
Estonia	Estonia	
Gabon		Gabon
Japan	Japan	
Kiribati		Kiribati
Lesotho		Lesotho
Malta	Malta	
Marshall Islands		Marshall Islands
Micronesia		Micronesia
Montenegro		Montenegro
Namibia		Namibia
Nauru		Nauru
Netherlands	Netherlands	
Palau		Palau
Peru		Peru
Portugal	Portugal	
Samoa		Samoa
Serbia		Serbia
South Africa		South Africa

It is noted that almost all Pacific Island States are included (Kiribati, Marshall Islands, Micronesia, Nauru, Palau and Samoa). Probably because their obesity level is already alarmingly high (from Nauru 61% is already obese). In these countries there is not much space left for additional rate of growth. It is noted that countries with a decreasing rate of growth comprise only 4% of the world population in 2016. This means that in 2016 about 96% of the world population is subjected to conditions that still result in an increasing rate of overweight and obesity prevalence. These findings confirm the observations as reported by the WHO (2021a) and by Shekar and Popkin (2020). Apparently, the factors that drive overweight and obesity prevalence affect all countries of the world. This supports the idea that country statistics can serve as a basis to investigate the significance and relative importance of factors in explaining overweight prevalence.

3.4 Variables included in regression

The dependent variable in the final regression model is prevalence of overweight. As motivated in *Section 3.3*, there is no added value of including prevalence of obesity as a second dependent variable.

Based on the literature review, in total 9 independent variables remain for inclusion in the current study. These variables are shown in *Table 3-2*, their sources are presented in *Table 3-3*.

Table 3-2: All independent variables included in regression analysis

Independent variable	Justification (refer to *Chapter 2*)
1. Adults (18+) physically inactive (%)	Hills, Andersen & Byrne (2011) Ladabaum, Mannalithara, Myer & Singh (2014)
2. Sugar available for consumption (kg/capita/year)	Hu, F. (2013) Stanhope (2016) Du, Tugendhaft, Erzse & Hofman (2018)
3. ln(Fat supply) (kg/capita/year)	Golay & Bobbioni (1997) Mensink, Zock, Kester & Katan (2003) Mozaffarian et al. (2004) Cohut (2018) Hu, S. et al. (2018)
4. Human Development Index (HDI)	Ataey, Jafarvand, Adham and Moradi-Asl (2020)
5. Urban population (% of total population)	Popkin (2019) Kuddus, Tynan and McBryde (2020) Tian, Zhao, Li, Wang and Shi (2014) Cooksey-Stowers, Schwartz and Brownell (2014) Ghosh-Dastidar et al. (2014)
6. KOF Globalization Index (total)	Costa-Font & Mas (2016)
7. KOF Globalization Index (economical dimension)	Costa-Font & Mas (2016)
8. KOF Globalization Index (social dimension)	Costa-Font & Mas (2016)
9. KOF Globalization Index (political dimension)	Costa-Font & Mas (2016)

Table 3-3: Sources of all variables included in regression analysis

Independent variable[1)]	Year of observation[2)]	Data source[3)]
Prevalence of overweight among adults, BMI ≥ 25 (%)[4)]	2016	(WHO, 2021c)
Prevalence of insufficient physical activity among adults 18+ years	2016	(WHO, 2021e)
Sugar available for consumption (kg/capita/year)	2016	(FAOSTAT, 2021a)
Fat supply (kg/capita/year)	2016	(FAOSTAT, 2021b)
Human Development Index (HDI)	2016	(UNDP, 2021)
Urban population (% of total population)	2016	(World Bank, 2021b)
KOF Globalization Index (total)	2016	(KOF, 2022)
KOF Globalization Index (economical dimension)	2016	(KOF, 2022)
KOF Globalization Index (social dimension)	2016	(KOF, 2022)
KOF Globalization Index (political dimension)	2016	(KOF, 2022)

Notes:
1) The variables are retrieved from databases as referred to in the column “Data source”.
2) The year of observation represents the year for which the data was published in the database.
3) The complete references of the data sources can be found in the *References* section.
4) Dependent variable of this study.

As motivated in *Chapter 2* the variables stunting, alcohol consumption, fast food restaurants per million inhabitants and GDP are excluded from the analysis. The complete set of variables discussed in the literature review is presented in *Table B - 1* of Appendix B, and the sources in *Table B - 2* of Appendix B.

The characteristics of the only dependent and the nine independent variables resulting from the literature review are discussed below. The descriptive statistics are given in *Table 3-4*. Similar details of the variables that are not included anymore, for reasons discussed above, can still be found in Appendix B.

Prevalence of overweight

The prevalence of overweight presents the percentage of adults aged 18 years and older of a country's population with a BMI≥25 (age-standardized estimate). BMI is calculated by dividing the persons weight by its height squared. The data are retrieved from the WHO database for the year 2016 (WHO, 2021c). The dataset from the WHO database originates originally from NCD Risk Factor Collaboration (NCD-RisC), which published their results in The Lancet of October 2017 (NCD Risk Factor Collaboration, 2017). The prevalence of overweight is obtained from 2416 population-based studies with measurements of height and weight on 97.4 million participants aged 20 years and older.

Physical inactivity

Physical inactivity presents the prevalence of insufficient physical activity among adults aged 18 years (age-standardized estimate). Physical inactivity is defined as the percentage of the county's population adults, aged 18 years and older, of a country attaining less than 150 minutes of moderate-intensity physical activity per week, or less than 75 minutes of vigorous-intensity physical activity per week, or equivalent. The data are retrieved from the WHO database for the year 2016 (WHO, 2021e). The data are estimated based on self-reported physical activity captured in questionnaire among the population covering activity at work/in the household, for transport, and during leisure time. If a country has no data available it is not included in the database.

Sugar supply

Sugar supply or sugar available for consumption is expressed in raw sugar equivalent, i.e. the amount of sugar of any form converted back to its equivalent unprocessed amount. The sugar forms included are cane, beet (raw and centrifugal), refined, confectionery and flavored. The data are retrieved from the FAOSTAT database for the year 2016 (FAOSTAT, 2021a). The sugar available for consumption is not a direct measurement of the actual consumption, but calculated from the quantity on stocks + quantities imported and produced – quantities exported, seed, fed to animals, waste and other non-food uses – ending stocks. The amount is expressed in kg/capita/year for each country included in the database. The food balance sheets, wherefrom the data are obtained, are derived statistics, of which the reliability depends of the underlying basic statistics of population, supply and utilization of foods and/or of their nutritive value. Often the data are supported by country's own expert opinions. Quality and coverage vary across countries and commodities

Fat supply

Fat supply presents the total national fat available for consumption as a 3-year average (2015-2017). Fat supply data are retrieved from FAOSTAT for the year 2016 (FAOSTAT, 2021b). Fat supply is not a direct measurement of the actual consumption, but calculated from the quantity on stocks + quantities imported and produced – quantities exported, seed, fed to animals, waste and other non-food uses – ending stocks. The amount is expressed in kg/capita/year for each country included in the database.

The fat supply data are retrieved from the so-called Suite of Food Security Indicators, which were established following the recommendation of experts in the World Food Security Committee (CFS). Data are obtained by expert judgment and in part on data, provided that sufficient data coverage is present to enable comparisons across regions and over time. Fat supply has been transformed into the natural logarithm of fat supply to obtain a more linear relationship between prevalence of overweight and fat supply.

Human Development Index (HDI)

The Human Development Index is a summary measure of key dimensions of human development and consists of three dimensions. The health dimension is measured as life expectancy at birth. The education dimension is measured as the mean years of schooling of adults aged 25 and older and the expected years of schooling of children entering school. Finally, the standard of living dimension is measured by using the logarithm of the gross national income (GNI), to reflect the diminishing importance of income. The Human Development Index is retrieved from the United Nations Human Development Data Center for the year 2016 (UNDP, 2021). The HDI is calculated using the geometric mean of the normalized indices of each of the three dimensions.

Urban population

Urban population is defined as the percentage of a country's total de facto population, which counts all residents regardless of legal status or citizenship, living in the urban areas. The data are retrieved from the World Bank database representing midyear estimates of the year 2016 (World Bank, 2021b). The data are defined by the national statistical offices and are collected and smoothed by the United Nations Population Division.

KOF Globalization Indices

KOF stands for 'Konjunkturforschungsstelle', which translates to Business Cycle Research Centre in English. The KOF Globalization Index (Haelg, 2019) is a composite index that measures globalization along the economic, social and political dimension for almost every country in the world on a scale of 1 (least) to 100 (most globalized). The Globalization indices are calculated on a yearly basis starting from 1970 (KOF, 2022). The data are retrieved form the KOF Swiss Economic Institute (KOF, 2021) for the year 2016. It is based on 43 different variables collected from secondary sources, the different variables can be found in the clarification of the KOF structure. Missing observations are linearly interpolated. It is noted that the KOF Globalization Indices include the economic index, the social index and the political index as well as the total index that includes all (*Table 3-2*, items 6 to 9 included).

An initial analysis is done with all variables of *Table 3-2* included. The initial analysis includes all 128 countries, developing and developed countries together, for which data is available.

The analysis is a so-called complete case analysis, as it excludes all the countries that miss data in one or more of the factors that were selected for inclusion in the analysis (*Table 3-2*, *Table 3-3*). The maximum number of variables that can be included should not exceed 8, based on a minimum preferred ratio between observations and variables of 15 (*Section 3.2*). This number reduces to 6 for the group developing countries only. Unless the data are imputed (Appendix C). However, from the literatures study more than 8 variables have been identified and are candidates for inclusion. So, some variables (factors) must be excluded from the analysis to still obtain reliable regression results. This may be possible if two variables are highly correlated, introducing a too high risk of collinearity. To check, the pairwise correlation are presented in the correlation matrix *Table A - 1* of Appendix A. In this matrix the Pearson's r values are shown for all possible pairs of variables. The pairwise correlation plot, showing the pairwise correlation graphically, is given in *Figure A - 1* of Appendix A (the addition 'of Appendix A' will be omitted, where it is noted that a letter in the reference refers to a corresponding appendix).

It can be observed that the (total) KOF Globalization Index correlates well with the HDI (Pearson's r >0.84). Also the KOF Social Dimension shows the highest correlation (Pearson's r >0.93), also with HDI. It is decided to omit both, the (total) KOF Globalization Index as well as the KOF Social Dimension from further analysis. The KOF Globalization Index is already presented by its dimensions, meaning that there is no need to also add it. The KOF Social Dimension is omitted because of its high correlation with HDI, where the health and education component of the HDI, are valued higher than the components collected in the KOF Social Dimension (Appendix D). For example McDonalds restaurants are already covered by sugar and fat supply, and IKEA can only have a market share if the degree of development has reached a certain minimum level. This means that the regression analysis will start with 7 independent variables and 128 countries if all countries with complete data are included, and 96 countries only developing countries are included. The regression analyses and results are presented in *Chapter 4*.

Table 3-4: Descriptive statistics all variables included in regression

Variable	Observations	Mean	Std. Dev.	Min	Max
Adults (18+) with BMI≥25 2016 (%)	189	48.6	16.8	18.3	88.5
Adults (18+) physically inactive (%)	159	28.4	10.9	5.5	67.0
Sugar available for consumption (kg/capita/year)	154	25.1	12.5	3.3	52.1
ln(Fat supply)	167	3.4	0.4	2.1	4.1
Human Development Index (HDI)	184	0.8	0.2	0.4	1.0
Urban population (% of total population)	188	58.1	22.7	12.4	100
KOF Globalization Index	186	61.9	14.5	30.3	90.7
KOF Economic Globalization Index	183	57.5	16.5	26.6	93.9
KOF Social Globalization Index	186	63.2	17.1	26.9	91.8
KOF Political Globalization Index	186	65.6	21.1	11.0	97.9

Chapter 4 – Analyses

The analysis is run with STATA v15. The dependent variable is the prevalence of overweight, refer to *Section 3.3*. The 7 independent variables eventually selected for the analysis are:

- Physical inactivity (% of population being less physically active than norm);
- Sugar supply (kg/capita/year);
- Fat supply (kg/capita/year);
- Human development index (HDI);
- Urban population (% of total population living in urban areas);
- KOF economical dimension;
- KOF political dimension.

The initial total number of countries available for analysis is 189.

After exclusion of all countries that have no data available for all the 7 independent variables, 128 countries remain. This means that 61 countries are excluded, because at least one variable is missing, it the intension is too run a complete case analysis. The descriptive statistics, *Table 3-4*, show that 35 countries have no data for sugar supply, 22 countries have missing data for fat supply and 30 countries have missing data for physical inactivity. The 61 countries missing correspond with 6% of the world's population. This may not seem a very large number, but the missing countries cover 32% of the maximum sample size (32% of countries miss information on one or more of the factors that are considered likely to contribute to overweight prevalence).

Whether the complete case analysis causes problems depends on the reason why the data are missing and the amount of data missing. If the amount of data missing is very small, less than 5%, the impact is negligible (thus a total of 8 countries missing would not be a real problem). If more than 5% of the countries are missing for a specific reason, and if these countries do also affect overweight prevalence, the missing data will cause biased results. If the data are not missing for a specific reason, thus completely at random, the regression is underpowered, but not biased. This means that the data are less good in predicting a real effect, if present.

One way to investigate whether the data are missing completely at random, or completely not at random, or something in between, may be investigated by comparing the group countries that miss data with the group that miss no data (Lodder, 2014). This is done visually for the factors that miss the most data (see *Figure A - 5*, *Figure A - 6*, *Figure A* - 7 and *Figure A - 8*). No clear patterns could be identified, other than that a relatively high number of excluded countries plot highest in overweight prevalence (*Figure A - 5*). These countries are almost all part of the Pacific Island States. Because this group is relatively small compared to the total, the impact is probably small. Enough countries seem to be left, with enough data available to still predict reasonable research results.

It may be observed that the developed countries are almost complete, as only 4 countries are missing from 36 in total (11%). A relatively large amount of developing countries miss data, 57 countries out of 153 must be excluded from a complete case analysis (37%).

Mostly the countries that are excluded do not miss data on all factors. Mostly Information lacks for one or two of the factors only. Excluding countries for this reason, therefore implies that statistical interesting information is left out, while this information could still be of interest. It is decided to impute the missing data. The impact and effectiveness of imputation is investigated for the dataset under consideration using different imputation methods. This is done by taking the 128 countries that have data for all independent variables (factors) available as the 'true' reference. Data are removed from this set, at random, to obtain a base set. The base set contains 128 countries, but 30% of the countries miss one or more of the factors included. In the next step, the base set is imputed multiple times to generate a large number of complete sets. The imputes are analyzed and pooled. The final results, if the multiple imputation is successful should produce more or less the same test results as the reference set. The advantage being that the true reference set is known and thus can be used for the verification of applied imputation procedures (Appendix C).

From Appendix C it is concluded that imputation provides added value and is considered the preferred model, with predictive mean matching (PMM) as the preferred imputation method above Bayesian regression. Although the differences between PMM and Bayesian regression appeared to be very small, the literature recommends PPM as an all-around method with relatively small biased estimates providing a better performance than many other methods (van Buuren, 2018).

Two regressions will be done. One regression that includes all countries available, the other with the developing countries only. It is decided to include the developing countries as a separate group for reasons already explained in *Chapter 1*. The developing countries are a vulnerable group, contain the majority of the world population, and their overweight prevalence increases at a relatively fast and increasing rate compared to the developed countries. Although the level of overweight prevalence is still relatively low in the developing countries, if the current trend is not stopped, they will pass the developed countries in overweight prevalence and the impact will be huge. The developing countries especially will be hit hard, since they already must deal with multiple problems of their own, such as infectious diseases, undernutrition and childhood stunting, which is referred to as 'double burden'. In view of the above, the so-called model 1 includes all countries and model 2 includes the subset of developing countries only.

Table 4-1 shows the regression results for model 1 and the same model 1 with imputed data. In both models all countries are included.

Table 4-1: Regression tables model 1 complete case vs imputed

Dependent variable: Overweight Prevalence Adults (18+) BMI≥25 (%)	Model-1 Complete case analysis		Model-1 PMM Imputed Stata++)	
Adults (18+) physically inactive (%)	0.038 (0.086)	*0.656*	0.060 (0.099)	*0.543*
Sugar consumption (kg/capita/year)	0.286 (0.088)	*0.001******	0.411 (0.100)	*0.000******
ln(Fat supply) (kg/capita/year)	9.685 (3.542)	*0.007******	14.918 (3.736)	*0.000******
Human Development Index (HDI)	22.462 (11.487)	*0.053**	21.200 (11.939)	*0.078**
Urban population (% of total population)	0.268 (0.058)	*0.000******	0.119 (0.052)	*0.023****
KOF Index (economical dimension)	-0.063 (0.070)	*0.371*	-0.085 (0.076)	*0.266*
KOF Index (political dimension)	-0.137 (0.057)	*0.017****	-0.185 (0.044)	*0.000******
Constant	-12.128 (7.210)	*0.095**	-18.426 (7.744)	*0.019****
Number of variables K and ratio N/K	7 [18.3]		7 [27.0]	
Number of observations N	128		189	
R^2 +)	0.741		0.644	
Adj. R^2	0.726		0.630	
Countries included	Developed and developing		Developed and developing	

Standard errors in parentheses. *P-values in italics, * p < 0.10, ** p < 0.05, *** p < 0.01*
+) The F-test complete case produced the following results: F(7, 120) = 49.07, p-value = 0.000
The F-test PMM imputed produced the following results: F(7, 177.3) = 37.74, p-value = 0.000
++) Stata and R-studio produce the same PMM imputation results, the reported numbers result from Stata

The complete case for model 1, did (just) not pass the Shapiro-Wilkins test for normality of the residuals (p-value = 0.035). It is noted that normality of the residuals is not one of the Gauss-Markov assumptions. This means that the estimators are still the best linear unbiased estimators. Prediction may not be that accurate, because confidence intervals are based on the assumption that the residuals are normally distributed. But, because the ratio between the number of countries versus the number of factors (independent variables) is quite large (18), it is considered that the regression produces still reliable results. The Gauss-Markov assumptions are checked next (BLUE). The model is linear in parameters and the sample is a random sample derived from trusted statistical databases, refer to *Table 3-3.* Heteroscedasticity is checked using the Breusch-Pagan/Cook-Weisberg test. The p-value is 0.77, indicating that heteroscedasticity is not present. The variance among the residuals is constant, refer also to *Figure A - 3*. That collinearity is no problem can be observed from the pairwise correlation matrices and the plots produced in Appendix A. The VIFs for the complete case analysis are about 5 or lower for the individual values and the average VIF is below 3

(*Table A - 3*), which confirms that collinearity is no problem. The imputed model 1 passed all test without any problems.

The multiple linear regression analysis is used to check if factors identified by literature are significant in explaining overweight prevalence using country statistics, and if yes, whether it would be possible to order these factors to their relative importance.

The fitted imputed regression model is:

$$Overweight = -18.426 + 0.060\, physical + 0.411\, sugar + 14.918\, fat + 21.200\, hdi + 0.119\, urban - 0.085\, kofecon - 0.185\, kofpol$$

With:

Overweight	= Overweight Prevalence Adults (18+) BMI≥25 (% country population)
Physical	= Adults (18+) physically inactive (% country population)
Sugar	= Sugar consumption (kg/capita/year)
Fat	= Natural logarithm of fat supply (kg/capita/year)
Hdi	= Human Development Index (HDI) (-)
Urban	= Urban population (% of country population)
Kofecon	= KOF Index (economical dimension) (-)
Kofpol	= KOF Index (political dimension) (-)

The imputed model 1 can explain 64% of the total variance in overweight prevalence with $R^2 = 0.64$, $F(7, 177.3) = 35.74$, and p-value = 0.000. The p-value of the F-test is significant at the 1% level, indicating that the regression model fits the data better than the model without independent variables. Adding independent variables improves the fit of the model. The regression is significant and the factors included in the model explain a large part of the variance. The multiple linear regression can be used to investigate the relative importance of the factors included.

All formulated hypotheses are checked and the results are summarized in Table 4-2.

Table 4-2: Results considered in relation to the research hypotheses (model 1)

No:	Hypotheses on factors identified in literature [1]	Accepted/ rejected
1A	Sugar and fat supply are positively associated with overweight prevalence	Accepted
1B	Fat supply dominates sugar supply in explaining changes in overweight prevalence	Rejected[5]
2	Urbanization is positively associated with overweight prevalence	Accepted
3A	Physical inactivity is positively associated with overweight prevalence.	Rejected[2]
3B	Overweight is more likely caused by excess calorie intake rather than increased physical inactivity.	Undetermined[3]
4	Globalization through the KOF globalization indices is positively associated with overweight prevalence	Rejected[4]
5A	Human development index is positively associated with overweight prevalence	Accepted
5B	Human development index and KOF globalization indices are well correlated.	Accepted
6	The relationship obtained by multiple linear regression is significant	Accepted

Notes:
1) Refer to *Chapter 2* and *Table 2-1*.
2) Hypothesis 3A: coefficient is positive but not significant.
3) A meaningful comparison was not possible because physical activity was insignificant.
4) The overall and social dimension of the KOF globalization index were excluded from the research. The economical dimension was not significant. The economical dimension and the political dimension both unexpectedly show a negative coefficient.
5) Fat supply dominates in the dominance analysis. However, fat supply is included as a transformed variable. The natural logarithm of the fat supply is used. This means that 1 unit raise corresponds with a 2.7 increase in terms of the untransformed variable. This explains the dominance, where it is noted that e^1 equals to 2.718. However, both fat supply and sugar supply always score high in terms of relative importance (higher than the other factors included).

The relatively low impact of the globalization index was not expected and this is therefore investigated in more detail. A sensitivity study is carried out comparing model 1 (complete case analysis) for different scenarios (Appendix E). In a first sensitivity analysis all KOF indices are included (social, economic, and political dimension) relative to the original. In a second sensitivity analysis, HDI is replaced by the KOF social dimension (*Table E - 1*) relative to the original. In a third sensitivity analysis lnfat is removed compared to the original, in the fourth and last sensitivity analysis lnfat is removed and HDI is replaced by KOF social dimension relative to the original (*Table E - 2*). The HDI and KOF social dimension are, to some extent interchangeable without affecting the results too much. They showed similar behavior and produced very similar trends in the regression (Appendix E). However, with HDI included instead of KOF social dimension better overall regression results are obtained (a higher R^2 and more of the dominant variables significant). It can be concluded that the KOF social dimension is the best of the globalization dimensions in explaining differences in overweight prevalence, which confirms the findings of Costa-Font and Mas (2016) who found in their study that the social globalization dimension provided a robust explanation for the increase of the obese and overweight population and the rise of calorie consumption, although on a limited number of countries (26) for reasons of data availability. The current research uses a more complete set of countries (189), covering all parts of the world and shows that HDI is a good alternative for the KOF social dimension.

Table 4-3 shows the relative importance of the factors that are included. The factors that are considered linked with globalization, being human development index, urbanization and the KOF political dimension explain 36.9% (18.7+13.7+4.5) of the total variance. Factors that matter the most according to the dominance analysis are linked to the calorie intake. Sugar supply and fat supply explain 45.7% of the total variance (21.6+24.1) and are significant with p-values of 0.000 and 0.000 respectively. Fat scores dominates in the dominance analysis. However, fat supply is included as a transformed variable. The natural logarithm of the fat supply is used. This means that 1 unit raise corresponds with a 2.7 increase in terms of the untransformed variable. This explains the dominance, where it is noted that (e^1) equals to 2.718, with (e) representing the base number of the natural logarithm. If the impact of for example a 500 kcal increase is investigated, sugar supply has more impact on overweight than fat supply, despite its relatively high energy density. However, both fat supply and sugar supply always score high in terms of relative importance compared to the other factors included.

Table 4-3: Dominance analysis model 1 complete case vs imputed

Variables	Model-1 Complete Case analysis		Model-1 PMM Imputed	
	Standardized Dominance Statistic (%)	Ranking	Standardized Dominance Statistic (%)	Ranking
Adults (18+) physically inactive (%)	7.7	6	8.5	6
Sugar consumption (kg/capita/year)	17.0	4	21.6	2
ln(Fat supply) (kg/capita/year)	19.4	2	24.1	1
Human Development Index (HDI)	19.1	3	18.7	3
Urban population (% of total population)	25.7	1	13.7	4
KOF Index (economical dimension)	7.9	5	8.8	5
KOF Index (political dimension)	3.4	7	4.5	7
Number of observations	128		189	
R^2	0.74		0.64	

The regression results for the group developing countries only (model 2) are summarized in *Table 4-4*, *Table 4-5* and *Table 4-6*. The comparison between both groups, model 1 versus model 2 is presented in *Table 4-7*.

Table 4-4: Regression tables model 2 complete case vs imputed

Dependent variable: Overweight Prevalence Adults (18+) BMI≥25 (%)	Model-2 Complete Case		Model-2 PMM Imputed Stata[++)]	
Adults (18+) physically inactive (%)	-0.011 (0.091)	*0.908*	0.035 (0.108)	*0.749*
Sugar consumption (kg/capita/year)	0.205 (0.106)	*0.056**	0.340 (0.120)	*0.006****
ln(Fat supply) (kg/capita/year)	7.211 (3.723)	*0.056**	15.043 (4.026)	*0.000****
Human Development Index (HDI)	33.355 (11.664)	*0.005****	28.879 (12.383)	*0.021***
Urban population (% of total population)	0.349 (0.063)	*0.000****	0.179 (0.056)	*0.002****
KOF Index (economical dimension)	-0.024 (0.078)	*0.756*	-0.025 (0.087)	*0.774*
KOF Index (political dimension)	-0.078 (0.058)	*0.184*	-0.161 (0.048)	*0.001****
Constant	-17.970 (9.143)	*0.053**	-28.460 (10.039)	*0.006****
Number of variables K and ratio N/K	7 [13.7]		7 [27.0]	
Number of observations N	96		153	
R^2 [+)]	0.782		0.668	
Adj. R^2	0.765		0.652	
Countries included	Developing		Developing	

Standard errors in parentheses. *P-values in italics, * p < 0.10, ** p < 0.05, *** p < 0.01*
[+)] The F-test complete case produced the following results: F(7, 88) = 45.18, p-value = 0.000
The F-test PMM imputed produced the following results: F(7, 141.8) = 34.17, p-value = 0.000
[++)] Stata and R-studio produce the same PMM imputation results, the reported numbers result from Stata

Comparison between imputed and the non-imputed (complete case) analysis again does not show dramatic differences, but there are differences in the coefficients. The analysis with the imputed model is considered to provide the most reliable results, as it includes more countries and uses all available information. It is considered to provide better estimates of the coefficients, although at cost of somewhat more uncertainty that likely results from the additional uncertainty introduced by the imputation process.

Table 4-5: Dominance analysis model 2 complete case vs imputed

Variables	Model-2 Complete Case		Model-2 PMM Imputed	
	Standardized Dominance Statistic (%)	Ranking	Standardized Dominance Statistic (%)	Ranking
Adults (18+) physically inactive (%)	7.2	5	7.5	6
Sugar consumption (kg/capita/year)	15.6	4	18.5	3
ln(Fat supply) (kg/capita/year)	18.6	3	26.3	1
Human Development Index (HDI)	21.6	2	21.4	2
Urban population (% of total population)	30.4	1	15.5	4
KOF Index (economical dimension)	5.3	6	7.6	5
KOF Index (political dimension)	1.3	7	3.2	7
Number of observations	96		153	
R^2	0.78		0.67	

The PMM imputed model includes more data than the complete case analysis. It can be observed that the ranking changed a bit. Calorie intake related parameters have become more dominant, at the cost of urbanization. This is more in line with some findings in literature that overweight grows faster in rural areas compared to urban area (Popkin, 2019). This would for the larger part be caused by a changing food environment and modernization resulting in lifestyle changes. The results in relation to the research hypotheses has been summarized also for the group developing countries only *Table 4-6*, where it has to be noted that the results are exactly similar for the group with all countries included (developed and developing). This in itself is also an interesting observation because some authors suggested that taking the subset of developing countries only as a single group may affect the research outcome, for example, relating to urbanization (Popkin, 2019).

Table 4-6: Results considered in relation to the research hypotheses (model 2)

No:	Hypotheses on factors identified in literature [1)]	Accepted/ rejected
1A	Sugar and fat supply are positively associated with overweight prevalence	Accepted
1B	Fat supply dominates sugar supply in explaining changes in overweight prevalence	Rejected[5)]
2	Urbanization is positively associated with overweight prevalence	Accepted
3A	Physical inactivity is positively associated with overweight prevalence.	Rejected[2)]
3B	Overweight is more likely caused by excess calorie intake rather than increased physical inactivity.	Undetermined[3)]
4	Globalization through the KOF globalization indices is positively associated with overweight prevalence	Rejected[4)]
5A	Human development index is positively associated with overweight prevalence	Accepted
5B	Human development index and KOF globalization indices are well correlated.	Accepted
6	The relationship obtained by multiple linear regression is significant	Accepted

Notes:
1) Refer to *Chapter 2* and *Table 2-1*.
2) Hypothesis 3A: coefficient is positive but not significant.
3) A meaningful comparison was not possible because physical activity was insignificant.
4) The overall and social dimension of the KOF globalization index were excluded from the research. The economical dimension was not significant. The economical dimension and the political dimension both unexpectedly show a negative coefficient.
5) Fat supply dominates in the dominance analysis. However, fat supply is included as a transformed variable. The natural logarithm of the fat supply is used. This means that 1 unit raise corresponds with a 2.7 increase in terms of the untransformed variable. This explains the dominance, where it is noted that e^1 equals to 2.718. However, both fat supply and sugar supply always score high in terms of relative importance (higher than the other factors included).

Table 4-7: Summary table models 1 and 2 imputed

Dependent variable: Overweight Prevalence Adults (18+) BMI≥25 (%)	Model-1 PMM Imputed			Model-2 PMM Imputed		
	Results		Rank	Results		Rank
Adults (18+) physically inactive (%)	0.060 (0.099)	*0.543*	6	0.039 (0.109)	*0.724*	*6*
Sugar consumption (kg/capita/year)	0.411 (0.100)	*0.000****	2	0.328 (0.124)	*0.010****	*3*
ln(Fat supply) (kg/capita/year)	14.918 (3.736)	*0.000****	1	15.431 (4.067)	*0.000****	*1*
Human Development Index (HDI)	21.200 (11.939)	*0.078**	3	28.809 (12.648)	*0.025***	*2*
Urban population (% of total population)	0.119 (0.052)	*0.023***	4	0.177 (0.056)	*0.002****	*4*
KOF Index (economical dimension)	-0.085 (0.076)	*0.266*	5	-0.025 (0.088)	*0.777*	*5*
KOF Index (political dimension)	-0.185 (0.044)	*0.000****	7	-0.165 (0.048)	*0.001****	*7*
Constant	-18.426 (7.744)	*0.019***		-29.219 (10.012)	*0.004****	
Number of variables K and ratio N/K	7 [27.0]			7 [27.0]		
Number of observations N	189			153		
R^2	0.644			0.669		
Countries included	Developed and developing			Developing		

Standard errors in parentheses. *P-values in italics, * p < 0.10, ** p < 0.05, *** p < 0.01*

The comparison between model 1 (all countries included) and model 2 (developed and developing countries included) produce similar results. The current research results clearly indicates the impact of globalization through HDI and urbanization, and the impact of calorie intake on overweight prevalence through sugar supply and fat supply. The results above show that there is no reason to consider that the group with developing countries included only would react differently on the factors causing overweight than the group with all countries included (which also includes the developed countries). This also confirms the worldwide trend of increasing overweight and obesity prevalence over time (*Section 3.3*), regardless of whether countries are developed or not.

Chapter 5 – Conclusions

The number of overweight and obese people increase rapidly in all countries of the world, with potentially severe consequences for society if the current trend is not stopped. Overweight and obesity are known causes for diabetes, cardiovascular disease, cancer and other illnesses like liver disease, infection, and depression. The associated trend of increasing healthcare expenditure is recognized all over the world, just as the negative indirect economic impact due to lost productivity and foregone economic growth because of lost workdays, lower productivity at work, mortality, and permanent disability.

The causes of overweight and obesity are well understood, overweight and obesity result from a sustained positive energy balance over a longer period, which is caused by consuming an excess of calories compared to the amount of calories expended.

In literature many factors have been identified that are likely to distort the energy balance. Conclusions on the relevance and importance of these factors in causing overweight and obesity are unfortunately often contradicting. In addition, these factors are often difficult to compare, because they are investigated on different populations, using different methods of analysis and it is not always clear how missing data were dealt with.

The current study therefore investigates the relevance of these factors in explaining overweight prevalence using statistical data collected at a national level including data from all countries of the world. It is investigated whether these factors are significant in explaining differences in overweight and obesity prevalence and if it is possible to investigate the relative importance of these factors (if significant).

As a first step the worldwide trends in overweight and obesity prevalence, as reported by international organizations such as the World Health Organization and the World Bank, were confirmed and investigated, using own data extracted from various trusted databases. The trends were checked for anomalous behaviors that might affect the current research. Japan, Singapore and South-Korea scored remarkably low in overweight prevalence despite a high level of development. This is explained dietary and cultural factors (*Section 3.3*). Another anomaly is the extremely high score in overweight and obesity of almost all Pacific Island States. This is explained by a rare combination of dependency of processed food from abroad (*Section 3.3*).

In general, the reported trends in overweight and obesity prevalence could be well reproduced using datasets for the years 2001, 2006, 2011 and 2016. The overweight and obesity prevalence grows in almost all countries of the world at an increasing rate, with only a few exceptions. Only in a limited number of countries, the overweight and obesity prevalence showed a small decrease in growth rate. For most of these countries, this could be explained by an already high overweight and obesity prevalence, not leaving much space left for additional growth (*Section 3.3*).

Numbers confirm that the growth in overweight and obesity in developing countries occurs faster than in developed countries, and that this trend is increasing over time. Overweight and obesity are strongly correlated, with a coefficient of determination close to 1 (*Figure 3-1*). Countries that have a large overweight population, have a relatively large number of obese people. For example, if about 30% of the people are overweight, about one third of this 30% is obese, if 70% of the people are overweight, about half of this 70% is obese. The method of measurement to determine overweight or obesity prevalence does not differ (only the cut-off point differs). In view of the above, and also for the simple reasons that prevention is to be preferred over treatment, overweight is selected as the dependent variable. There is still space for prevention, because a the majority of the world population can still be saved from becoming overweight and obese. They did not reach that level yet.

Different factors were selected from literature that potentially can affect overweight prevalence. Their relevance was investigated by multiple linear regression with overweight as the dependent variable. The overall regression at least some of the identified factors can explain differences in overweight prevalence. This already confirms the research question. It is possible to identify factors that are significant in explaining differences in overweight prevalence. If more than one factor is significant, than it will be possible too to conclude on their relative importance.

From the literature review the following variables were identified for potential inclusion in the study: physical inactivity, sugar supply, fat supply, human development index, urbanization and the KOF indices (namely the social, economic and political dimensions). From these parameters a subset is selected for inclusion in the current study. Some variables are excluded from the analysis, because they cannot provide sufficient added value over the already included variables. For example, the total KOF Globalization Index is excluded because it is a container variable which already includes the social, economic and political dimensions. It makes much more sense to include the KOF dimensions separately to enable the checking for their relevance and relative importance, rather than also including the container variable. For the same reason, food supply is excluded. Food supply is a container variable, which already includes the variables fat supply and sugar supply. Gross domestic product is excluded, because the economic dimension is also included in the human development index (via gross national income) and/or in the KOF economic dimension. Some variables are excluded because it is impossible to compare their impact in a meaningful way over different countries. For example, alcohol consumption will have a different impact for different countries. For the reason that in countries with an Islamic background, alcohol consumption is restricted or prohibited. Some variables are not included because their impact can only be studied in the time domain, such as childhood stunting.

Eventually 7 independent variables are selected for inclusion in the current study. It is important that the number of observations, that is the number of countries, remains large enough to produce reliable results if all 7 independent variables are included. Therefore, to keep the group of countries large, but also to not unnecessarily exclude still available information in countries that otherwise would have been excluded, it is decided to impute missing data. The impact of imputing data on factors with missing data is investigated by a trial on an artificial dataset. From this trial it is considered that the imputed models provide the most complete and reliable outcome, although at the costs of a somewhat smaller coefficient of determination (Appendix C).

Different hypotheses were formulated (*Table 2-1*) and verified for the variables that eventually are included: physical inactivity, sugar supply, fat supply, human development index, urbanization, KOF political and KOF economical dimensions. Verification is done by multiple linear regression that is carried out on the group with all countries included (model 1), this group includes developing and developed countries. The verification is also done on a second group comprising of the developing countries only (model 2).

The variable included are able to explain between 65% and 80% of the total variance in overweight prevalence, depending on the number of countries included.

Based on the regression analyses, the hypotheses are approved or rejected (*Table 4-2*). The factors sugar supply and fat supply are significant in explaining overweight prevalence, with the model corrected for the other factors included, at the 1% significance level. This appeared the case for all imputed models (model 1 and 2). The factors human development index and urbanization are significant at the 10% and 5% level respectively for the imputed model with all countries included (model 1), and at the 5% and 1% respectively for the imputed model with developing countries only included (model 2).

The included factors associated with calorie intake, fat and sugar supply, contribute in all imputed models more in explaining overweight prevalence, than all the other included factors together. Fat and sugar supply explain about 45% of the variance that could be attributed to the included variables in both models. The factors that are considered associated with globalization, namely human development index and urbanization, explain about 30% to 35% of the variance that can be attributed to the included variables (for the models 1 and 2 respectively). It is suggested that globalization creates a climate for a changing food environment and a change in lifestyle, resulting in an increased calorie intake. The change in lifestyle is frequently linked with lesser physical activity to explain increased overweight prevalence. In this study, however, physical inactivity could not be associated with overweight prevalence. The p-values are well above 0.5 for both the imputed models. A review of the method of measurement shows that the measurement of physical inactivity is subjective (self-reported). In addition, the threshold value used to call somebody physically inactive seems low (2.5 hours of medium intensity of activity each week is not that much in terms of calorie expenditure compared to the calorie intake of a hamburger for example). The method of measurement might be useful in well controlled surveys at a more region scale, but may not be that effective as measure to express energy expenditure at a country level.

The KOF economic dimensions, one of the KOF globalization indices, could also not be associated with overweight prevalence, regardless of whether human development index was included or not. The p-values were about 0.25 and about 0.75 for the imputed models 1 and 2 respectively. Globalization is still considered the driving factor behind a changing food environment, if expressed by urbanization and human development index (this study), eventually resulting in more fat and sugar consumption, but the results of the current study suggest that the KOF economic dimension is not the best globalization factor to explain differences in overweight prevalence. Instead, the results suggest that the human development index and urbanization are better in explaining overweight prevalence. These factors are considered linked to globalization because a high standard of living, a high life expectancy at birth, a high degree of education and as a high level of urbanization in a country, are all strongly related with international trade and economic and social interactions at a country overarching level. The KOF political index appeared significant at the 1% level, with a negative sign on its coefficient. This index refers to the country's ability to engage in international political cooperation. This negative sign cannot be explained. It is unclear why more political engagement of a country would result in less overweight prevalence. The opposite is considered more likely, a higher degree of political engagement, as a sign of increased globalization resulting in a higher degree of change of the food environment, and consequently an increase of processed food high in sugar and fat and thus an increase in overweight prevalence.

Chapter 6 – Discussion and recommendations

This cross sectional research is one of the fewer that could be found in literature that aimed to include information as complete as possible. All countries were included. However, in the complete case analyses countries had to be excluded for missing data on one or more of the independent variables (the factors under investigation).

To still obtain a group countries large enough to produce reliable results when all independent variables identified by literature (7 in total) are included data was imputed for factors that missed data. Most of the countries that were excluded for reasons of missing data on one or more of the factors, missed data on maximum 3 factors. Exclusion of these countries would mean that still available on the other 4 factors of excluded countries would not be accounted for.

A trial was carried out to check for the effectiveness of different imputation procedures. A complete set of data was available for 128 data. Data was randomly and artificially removed from factors in such a way that 90 countries remained (30% artificially missing). Imputation was carried out and the results were compared with the dataset before the removal of data. The result appeared very good for the imputation methods investigated. Apparently, the method works well, provided that data is missing at random, or in a way not affecting the expected outcome because they are missing. The missingness of data was investigated at depth. A relatively low amount of countries were excluded for the complete case analysis (4 out of 36 = 11%). This amount is not that large and the impact is considered negligibly small. A considerably larger amount of data was missing from the group developing countries (57 out of 153 = 37%). It was checked whether countries were missing for specific reason and whether the missingness were likely to affect the impact of factors on overweight prevalence. This appeared not the case. The countries with missing data were also well distributed over the full range of overweight prevalence investigated, with perhaps a relatively large amount of missing countries in case of the Pacific Island States. However, this is not considered to impact the overall result to a great extent, because the Pacific Island States reflect still a relatively small amount of countries of the total number of developing countries. The test results did show some changes between the complete case analyses and the imputed model analyses. In view of the above, it is considered that the imputed model explains the relative contribution of the different factors better than the complete case model. In view of the differences, and the potential impact of missing data it is recommended that this aspect is also well embedded in checking procedures of the research methodology. Thus in addition to the BLUE criteria and checks for normality of the residuals.

The practical implication of this study is that it is confirmed that country statistics can be of added value in identifying causes of overweight prevalence. Statistics obtained at national level can tell something about overweight prevalence of the world population. Because the people in different countries are subjected to the same factors, but to different degree, resulting in different overweight prevalence numbers. However, the results should be understood to reflect differences at the global scale and only in terms of average trends. No distinction is made towards regional differences or other differences that may exists within countries. For detailed answers on specific questions, specific research will still be required on smaller samples sizes.

Another practical implication of this research is that it confirms the extent of the overweight and obesity problem worldwide and confirms the importance of food policies to mitigate this problem with calorie intake clearly dominant over all other factors. Globalization is a factor that cannot easily be stopped, changed or guided. The continuing change of the food environment, closely linked with globalization and modernization, also confirmed by this research through the relative importance of the factors human development and urbanization, may and must be changed for the better. This includes everything that may be of impact on the too much and unhealthy calorie intake. This may include policies to force the food industry to start producing more healthy food; the priority should perhaps be on food quality (in terms of health) rather than on producing unhealthy food because it results in higher profits. Policies should take care that quality food comes within reach of all people of the world.

References

Ataey, A., Jafarvand, E., Adham, D. & Moradi-Asl, E. (2020). The Relationship Between Obesity, Overweight, and the Human Development Index in World Health Organization Eastern Mediterranean Region Countries. *Journal of Preventative Medicine and Public Health, 53*(2), 98-105.

Austin, P.C. & Steyerberg, E.W. (2015). The number of subjects per variable required in linear regression analyses. *Journal of Clinical Epidemiology, 68*(6), 627-636.

Azen, E. & Budescu, D.V. (2006). Comparing Predictors in Multivariate Regression Models: An Extension of Dominance Analysis. *Journal of Educational and Behavioral Statistics, 31*(2), 157-180.

Bowerman, B.L., O'Connell, R.T. & Murphree, E.S. (2011). Business Statistics In Practice (6^{th} edition). McGraw-Hill Irwin.

Burgoine, T., Sarkar, C., Webster, C.J. & Monsivais, P. (2018). Examining the interaction of fast-food outlet exposure and income on diet and obesity: evidence from 51,361 UK Biobank participants. *International Journal of Behavioral Nutrition and Physical Activity, 15*(1), 1-12.

Burki, T. (2021). European Commission classifies obesity as a chronic disease. *The Lancet Diabetes & Endocrinology, 9*(7), 418.

CDC. (2021a). Body Mass Index (BMI). Retrieved from: https://www.cdc.gov/healthyweight/assessing/bmi/index.html

CDC. (2021b). Adult Obesity Causes & Consequences. Retrieved from: https://www.cdc.gov/obesity/adult/causes.html

Cohut, M. (2018). Fats or carbs: What causes obesity? Retrieved from: https://www.medicalnewstoday.com/articles/322481

Convington, M. (2017). Why is Obesity a Disease? *Obesity Medicine Association.* Retrieved from: https://obesitymedicine.org/why-is-obesity-a-disease/

Cooksey-Stowers, K., Schwartz, M.B. & Brownell, K.D. (2017). Food Swamps Predict Obesity Rates Better Than Food Deserts in the United States. *International Journal of Environment Research and Public Health, 14*(11), 1366-1386.

Costa-Font, J. & Mas, N. (2016). 'Globesity'? The effects of globalization on obesity and caloric intake. *Food Policy, 64*, 121-132.

Davis, B. & Carpenter, C. (2009). Proximity to Fast-Food Restaurants to Schools and Adolescent Obesity. *American Journal of Public Health, 99*(3), 505-510.

Dhakal, C.K. & Khadka, S. (2021). Heterogeneities in Consumer Diet Quality and Health Outcomes of Consumer by Store Choice and Income. *Nutrients, 13*(4), 1046-1062.

Drewnowski, A. (2004). Obesity and the food environment: Dietary energy and diet costs. *American Journal of Preventive Medicine, 27*(3), 154-162.

Du, M., Tugendhaft, A., Erzse, A. & Hofman, K.J. (2018). Sugar-Sweetened Beverage Taxes: Industry Response and Tactics. *Yale Journal of Biology and Medicine, 91*(2), 185-190.

Encyclopedia Britannica. (2022). Definition Modernization. Retrieved from: https://www.britannica.com/topic/modernization

FAOSTAT. (2021a). Sugar (Raw Equivalent) – Item code 2542. Retrieved from: https://www.fao.org/faostat/en/#data/FBS

FAOSTAT. (2021b). Average fat supply (3-year average 2015-2017) – Item code 21061. Retrieved from: https://www.fao.org/faostat/en/#data/FS

FAOSTAT. (2021c). Food supply. Retrieved from: https://www.fao.org/faostat/en/#data/FBS

Fraser, L.K., Clarke, G.P., Cade, J.E. & Edwards, K.L. (2012). Fast Food and Obesity: A Spatial Analysis in a Large United Kingdom Population of Children Aged 13-15. *American Journal of Preventative Medicine 42*(5), e77-e85.

Garcia, G., Sunil, T.S. & Hinojosa, P. (2012). The Fast Food and Obesity Link: Consumption Patterns and Severity of Obesity. *Obesity Surgery, 22*(5), 810-818.

Golay, A. & Bobbioni, E. (1997). The role of dietary fat in obesity. *International Journal of Obesity and Related Metabolic Disorders, 21*(3), S2-11.

Ghosh-Dastidar, B., Cohen, D., Hunter, G., Zenk, S.N., Huang, C., Beckman, R. & Dubowitz, T. (2014). Distance to Store, Food Prices, and Obesity in Urban Food Deserts. *American Journal of Preventative Medicine, 47*(5), 587-595.

Green, S.B. (1991). How Many Subjects Does It Take To Do A Regression Analysis. *Multivariate Behavioral Research, 26*(3), 499-510.

Haelg, F. (2019). The KOF Globalisation Index – A Multidimensional Approach to Globalisation. *Journal of Economics and Statistics, 240*(5), 691–696.

Harvard School of Public Health. (n.d.). Types of Fat. Retrieved from: https://www.hsph.harvard.edu/nutritionsource/what-should-you-eat/fats-and-cholesterol/types-of-fat/

Hawley, N.L. & McGarvey, S.T. (2015). Obesity and Diabetes in Pacific Islanders: the Current Burden and the Need for Urgent Action. *Current Diabetes Reports, 29*, 1-10.

Hills, A.P., Andersen, L.B. & Byrne, N.M. (2011). Physical activity and obesity in children. *British Journal of Sports Medicine, 45*, 866-870.

Hu, F.B. (2013). Resolved: there is sufficient scientific evidence that decreasing sugar-sweetened beverage consumption will reduce the prevalence of obesity and obesity-related diseases. *Obesity Reviews, 14*(8), 606-619.

Hu, et al. (2018). Dietary Fat, but Not Protein or Carbohydrate, Regulates Energy Intake and Causes Adiposity in Mice. *Cell Metabolism, 28*(3), 415-431.e4.

IMF. (2018). World Economic Outlook: Challenges to Steady Growth. Washington, DC: IMF.

Jakobsen, J.C., Gluud, C., Wetterslev, J. & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Medical Research Methodology, 17*, 162-172.

Jeffery, R.W., Baxter, J., McGuire, M. & Linde, J. (2006). Are fast food restaurants an environmental risk factor for obesity? *International Journal of Behavioral Nutrition and Physical Activity, 3(2)*, 1-6.

KFC. (2021). World map: Number of fast food restaurants. Retrieved from: https://global.kfc.com/our-locations/

Kim, D.D. & Basu, A. (2016). Estimating the Medical Care Costs of Obesity in the United States: Systematic Review, Meta-Analysis, and Empirical Analysis. *Value in Health, 19*(5), 602-613.

Kinge, J.M., Strand, B.H., Vollset, S.E. & Skirbekk, V. (2015). Educational inequalities in obesity and gross domestic product: evidence from 70 countries. *Journal of Epidemiology and Community Health, 69*(12), 1141-1146.

KOF. (2022). KOF Globalisation Index. Retrieved from: https://kof.ethz.ch/en/forecasts-and-indicators/indicators/kof-globalisation-index.html

KOF. (2021). Globalisation Index: Structure, variables and weights. Retrieved from: https://ethz.ch/content/dam/ethz/special-interest/dual/kof-dam/documents/Globalization/2021/KOFGI_2021_structure.pdf

Kuddus, M.A., Tynan, E. & McBryde, E. (2020). Urbanization: a problem for the rich and the poor? *Public Health Reviews, 41*, 1-4.

Labine, J. (2020). Redefining obesity: New guideline shifts focus to treating excess fat as a chronic disease. *Edmonton Journal*. Retrieved from: https://edmontonjournal.com/health/diet-fitness/redefining-obesity-new-guideline-shifts-focus-to-treating-excess-fat-as-a-chronic-disease

Ladabaum, U., Mannalithara, A., Myer, P.A. & Singh, G. (2014). Obesity, Abdominal Obesity, Physical Activity, and Caloric Intake in US Adults: 1988 to 2010. *The American Journal of Medicine, 127*(8), 717-727.e12.

Lavery, M.R., Acharya, P., Sivo, S.A. & Xu, L. (2017). Number of predictors and multicollinearity: What are their effects on error and bias in regression? *Communications in Statistics – Simulation and Computation, 48*(1), 27-38.

Lodder, P. (2014). *Advising on research methods: Selected topics 2013*. To Impute or not Impute: That's the Question. Johannes van Kessel Publishing.

Malik, V.S. & Hu, F.B. (2019). Sugar-Sweetened Beverages and Cardiometabolic Health: An Update of the Evidence. *Nutrients, 11*(8), 1840-1857.

McCleary, L. (2002). Using multiple imputation for analysis of incomplete data in clinical research. *Nursing Research, 51*(2), 339-343.

McDonalds. (2021). Number of fast food restaurants. Retrieved from: https://corporate.mcdonalds.com/corpmcd/our-company/where-we-operate.html

Mensink, R.P., Zock, P.L., Kester, A.D.M. & Katan, M.B. (2003). Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. *American Journal of Clinical Nutrition, 77*(5), 1146-1155.

Merriam-Webster Dictionary. (2022). Definition Globalization. Retrieved from: https://www.merriam-webster.com/dictionary/globalization

Mozaffarian, D., Pischon, T., Hankinson, S.E., Rifai, N., Joshipura, K., Willett, W.C. & Rimm, E.B. (2004). Dietary intake of trans fatty acids and systemic inflammation in women. *American Journal of Clinical Nutrition, 79*(4), 6060-612.

Muhammad, H.F.L. (2018). Obesity as the Sequel of Childhood Stunting: Ghrelin and GHSR Gene Polymorphism Explained. *Acta Medica Indonesiana, 50*(2), 159-164.

Müller, M.J. & Geisler, C. (2017). Defining obesity as a disease. *European Journal of Clinical Nutrition, 71*, 1256-1258.

NCD Risk Factor Collaboration (NCD-RisC). (2017). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *The Lancet, 390*(10113), 2627-2642.

Ng, M. et al. (2014). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet, 384*(9945), 766-781.

Ng, S.W. & Popkin, B.M. (2012). Time use and physical activity a shift away from movement across the globe. *Obesity Reviews, 13*(8), 659-680.

NHS. (2019). Causes Obesity. Retrieved from: https://www.nhs.uk/conditions/obesity/causes/

OMA. (2013). AMA House of Delegates Adopts Policy to Recognize Obesity as a Disease. *Obesity Medicine Association.* Retrieved from: https://obesitymedicine.org/ama-adopts-policy-recognize-obesity-disease/

Popkin, B.M. (2019). Rural areas drive the global weight gain. *Nature, 569*, 200-201.

Reiff, N. (2020). 10 Largest Beverage Companies. Retrieved from: https://www.investopedia.com/news/worlds-largest-food-and-beverage-players-2017-nestle-pepsi-coke-topple-peers/

Rolls, B.J. (2017). Dietary energy density: Applying behavioural science to weight management. *Nutrition Bulletin, 42*(3), 246-253.

Romieu, I. et al. (2017). Energy balance and obesity: what are the main drivers? *Cancer Causes & Control, 28*(3), 247-258.

Ross, S.E., Flynn, J.I. & Pate, R.R. (2016). What is really causing the obesity epidemic? A review of reviews in children and adults. *Journal of Sports Sciences, 34*(12), 1148-1153.

Sayon-Orea, C., Martinez-Gonzalez, M.A. & Bes-Rastrollo, M. (2011). Alcohol consumption and body weight: a systematic review. *Nutrition Reviews, 69*(8), 419-431.

Schmidt, F.L. (1971). The Relative Efficiency of Regression and Simple Unit Predictor Weights in Applied Differential Psychology. *Educational and Psychological Measurement, 31*(3), 699-714.

Schneider, A., Hommel, G. & Blettner, M. (2010). Linear Regression Analysis Part 14 of a Series on Evaluation of Scientific Publications. *Deutsches Ärzteblatt International, 107*(44), 776-782.

Senauer, B. & Gemma, M. (2006). Why Is the Obesity Rate So Low in Japan and High in the U.S.? Some Possible Economic Explanations. Working Papers 14321. University of Minnesota, The Food Industry Center.

Sharma, A.M. & Campbell-Scherer, D.L. (2017). Redefining obesity: Beyond the numbers. *Obesity, 25*(4), 660-661.

Shekar, M. & Popkin, B. (2020). Obesity Health and Economic Consequences of an Impending Global Challenge. *Human Development Perspectives series.* Washington, DC: World Bank.

Silventoinen, K., Sans, S., Tolonen, H., Monterde, D., Kuulasmaa, K., Kesteloot, H. & Tuomilehto, J. (2004). Trends in obesity and energy supply in the WHO MONICA Project. *International Journal of Obesity, 28*, 710-718.

Smethers, A.D. & Rolls, B.J. (2018). Dietary Management of Obesity: Cornerstones of Healthy Eating Patterns. *Medical Clinics of North America, 102*(1), 107-124.

Stanhope, K.L. (2016). Sugar consumption, metabolic disease and obesity: The state of the controversy. *Critical Review in Clinical Laboratory Sciences, 53*(1), 1-34.

Stavseth, M.R., Clausen, T. & Røislien, J. (2018). How handling missing data may impact conclusions: A comparison of six different imputation methods for categorical questionnaire data. *SAGE Open Medicine, 7*, 1-12.

Subway. (2021). Number of fast food restaurants. Retrieved from: https://www.subway.com/en-us/exploreourworld

Swanbrow, D. (2012). A Tale of Turkey Tail: The part of the bird that's best left uneaten. Michigan News. University of Michigan. Retrieved from: https://news.umich.edu/a-tale-of-turkey-tail-the-part-of-the-bird-best-left-uneaten/

The Coca-Cola Company. (2021). How does the company address obesity? Retrieved from: https://www.coca-colacompany.com/faqs/how-does-coca-cola-address-obesity

Tian, X., Zhao, G., Li, Y., Wang, L. & Shi, Y. (2014). Overweight and Obesity Differences of Chinese Population Between Different Urbanization Levels. *The Journal of Rural Health, 30*(1), 101-112.

Traversy, G. & Chaput, J. (2015). Alcohol Consumption and Obesity: An Update. *Current Obesity Reports, 4*(1), 122-130.

Tremmel, M., Gerdtham, U., Nilsson, P.M. & Saha, S. (2017). Economic Burden of Obesity: A Systematic Literature Review. *International Journal of Environmental Research and Public Health, 14*(4), 1-18.

Tsugane, S. (2021). Why has Japan become the world's most long-lived country: insights form a food and nutrition perspective. *European Journal of Clinical Nutrition, 75*, 921-928.

UNDP. (2021). Human Development Index (HDI). Retrieved from: http://hdr.undp.org/en/indicators/137506

UNDP. (2022). United Nations Development Programme: Human Development Index (HDI). Retrieved from: http://hdr.undp.org/en/content/human-development-index-hdi

Van Buuren, S. (2018). *Flexible Imputation of Missing Data* (2nd ed.). Taylor & Francis Ltd.

Van der Merwe, M.T. & Pepper, M.S. (2006). Obesity in South Africa. *Obesity Reviews, 7*(4), 315-322.

WHO. (2003). Diet, food supply and obesity in the Pacific. Retrieved from: https://apps.who.int/iris/bitstream/handle/10665/206945/9290610441_eng

WHO. (2021a). Obesity and Overweight Key facts. Retrieved from: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

WHO. (2021b). Malnutrition. Retrieved from: https://www.who.int/health-topics/malnutrition#tab=tab_1

WHO. (2021c). Prevalence of overweight among adults, BMI >= 25 (age-standardized estimate) (%). Retrieved from: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-overweight-among-adults-bmi-=-25-(age-standardized-estimate)-(-)

WHO. (2021d). Prevalence of obesity among adults, BMI >= 30 (age-standardized estimate) (%). Retrieved from: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-obesity-among-adults-bmi-=-30-(age-standardized-estimate)-(-)

WHO. (2021e). Prevalence of insufficient physical activity among adults aged 18+ years (age-standardized estimate) (%). Retrieved from: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-insufficient-physical-activity-among-adults-aged-18-years-(age-standardized-estimate)-(-)

WHO. (2021f). Stunting prevalence among children under 5 years of age (% height-for-age <-2 SD). Retrieved from: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/gho-jme-country-children-aged-5-years-stunted-(-height-for-age--2-sd)

WHO. (2021g). Alcohol, total per capita (15+) consumption (in litres of pure alcohol). Retrieved from: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/total-(recorded-unrecorded)-alcohol-per-capita-(15-)-consumption

Wright, S.M. & Aronne, L.J. (2012). Causes of obesity. *Abdominal Radiology, 37*, 730-732.

World Bank. (2021a). GDP per capita (current US$). Retrieved from: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD

World Bank. (2021b). Urban population (% of total population). Retrieved from: https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS

World Bank. (2021c). Population (total). Retrieved from: https://data.worldbank.org/indicator/SP.POP.TOTL

Wyatt, S.B., Winters, K.P. & Dubbert, P.M. (2006). Overweight and Obesity: Prevalence, Consequence, and Causes of a Growing Public Health Problem. *The American Journal of the Medical Sciences, 331*(4), 166-174.

Zobel, E.H., Hansen, T.W., Rossing, P. & Scholten, B.J. (2016). Global Changes in Food Supply and the Obesity Epidemic. *Current Obesity Reports, 5*, 449-455.

Appendix A – Statistical tests

This page intentionally left blank.

Table A - 1: Pairwise correlation matrix all variables included in the regression

Variables	Adults (18+) BMI>25 (%)	Adults (18+) physically inactive (%)	Sugar consumption (kg/capita/year)	ln(Fat supply) (kg/capita/year)	Human Development Index (HDI)	Urban population (% of total population)	KOF Globalization Index (total)	KOF Globalization Index (economical dimension)	KOF Globalization Index (social dimension)	KOF Globalization Index (political dimension)
Adults (18+) BMI>25 (%)	1									
Adults (18+) physically inactive (%)	0.49	1								
Sugar consumption (kg/capita/year)	0.67	0.54	1							
ln(Fat supply) (kg/capita/year)	0.71	0.45	0.56	1						
Human Development Index (HDI)	0.67	0.49	0.61	0.83	1					
Urban population (% of total population)	0.59	0.52	0.60	0.65	0.70	1				
KOF Globalization Index (total)	0.48	0.34	0.58	0.77	0.84	0.63	1			
KOF Globalization Index (economical dimension)	0.51	0.28	0.53	0.68	0.76	0.53	0.83	1		
KOF Globalization Index (social dimension)	0.67	0.45	0.63	0.83	0.93	0.62	0.83	0.82	1	
KOF Globalization Index (political dimension)	0.09	0.15	0.32	0.40	0.40	0.39	0.73	0.27	0.29	1

Figure A - 1: Pairwise correlation plot all variables included in the regression

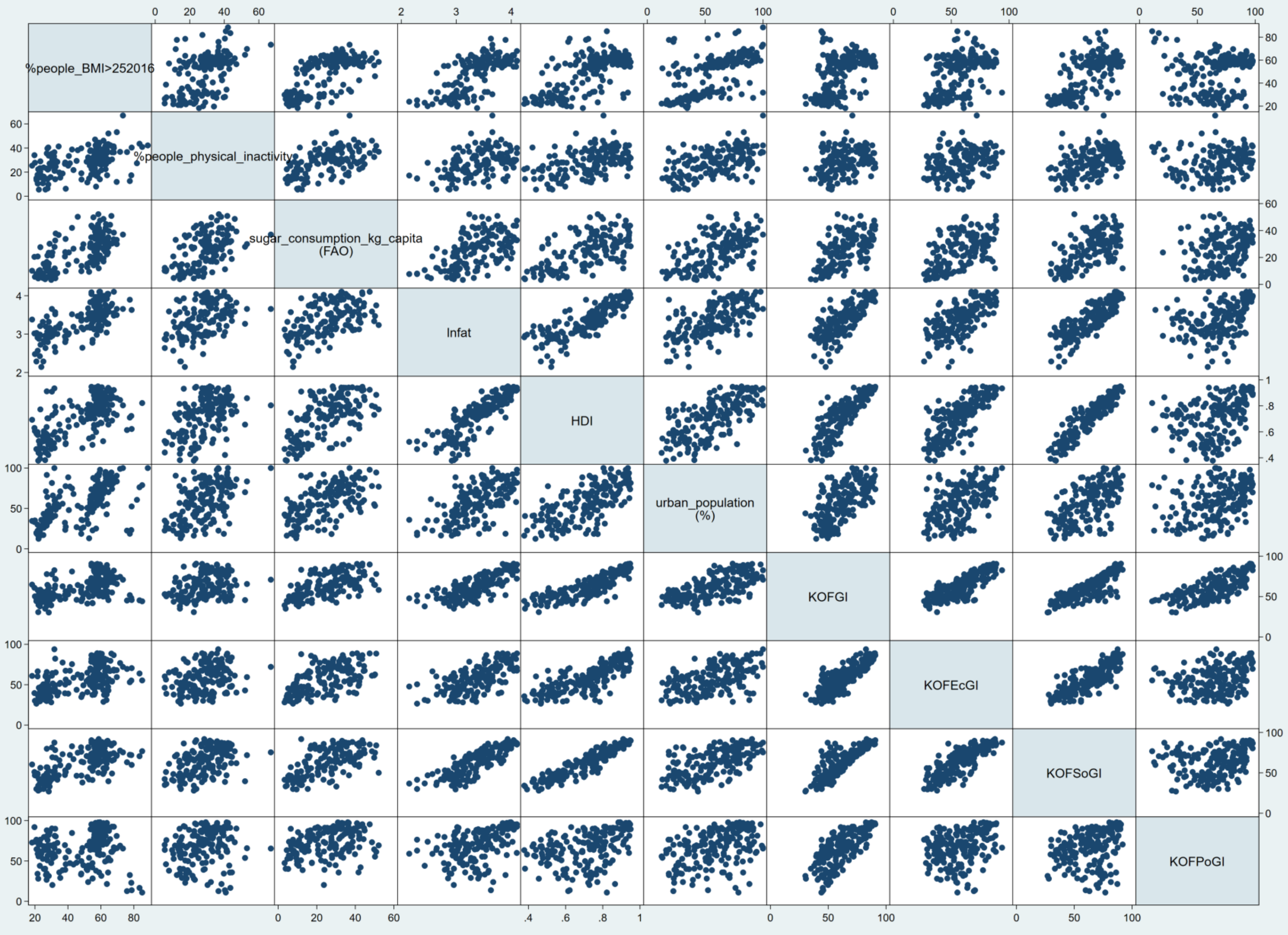
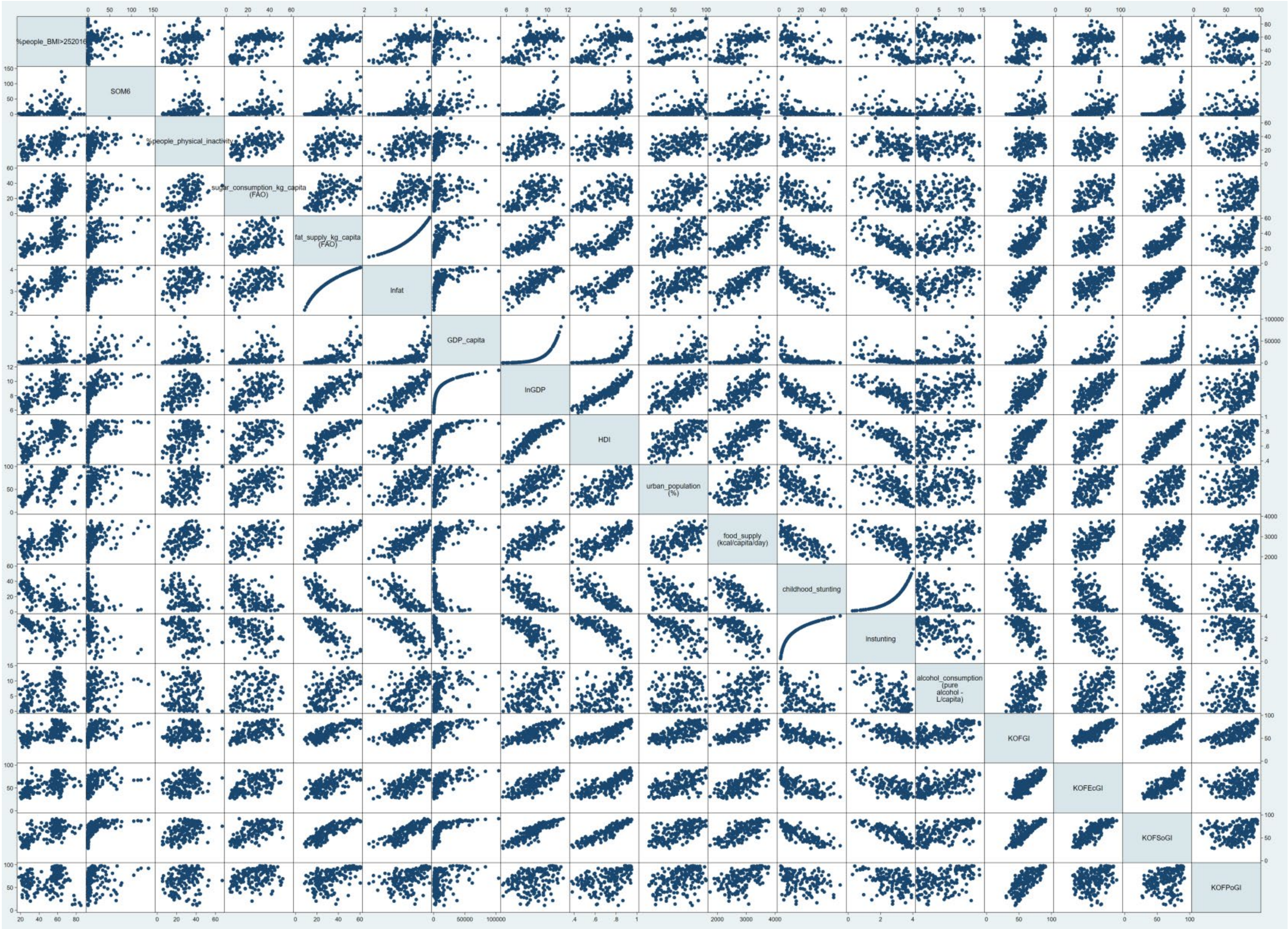


Table A - 2: Pairwise correlation matrix all variables cf. Table B - 1

Variables	Adults (18+) BMI≥25 (%)	Total no. fast food restaurants per 1 million inhabitants	Adults (18+) physically inactive (%)	Sugar consumption (kg/capita/year)	Fat supply (kg/capita/year)	ln(Fat supply) (kg/capita/year)	GDP per capita (current USD)	ln(GDP per capita) (current USD)	Human Development Index (HDI)	Urban population (% of total population)	Food supply (kcal/capita/day)	Childhood stunting (%)	ln(Childhood stunting) (%)	Pure alcohol consumption (L/capita/year)	KOF Index (total lumped sum)	KOF Index (total economic)	KOF Index (total social)	KOF Index (total political)
Adults (18+) BMI≥25 (%)	1																	
Total no. fast food restaurants per 1 million inhabitants	0.32	1																
Adults (18+) physically inactive (%)	0.49	0.37	1															
Sugar consumption (kg/capita/year)	0.67	0.41	0.54	1														
Fat supply (kg/capita/year)	0.67	0.54	0.41	0.52	1													
ln(Fat supply) (kg/capita/year)	0.71	0.51	0.45	0.56	0.97	1												
GDP per capita (current USD)	0.35	0.66	0.31	0.37	0.74	0.66	1											
ln(GDP per capita) (current USD)	0.61	0.65	0.53	0.61	0.84	0.83	0.82	1										
Human Development Index (HDI)	0.67	0.58	0.49	0.61	0.83	0.83	0.71	0.94	1									
Urban population (% of total population)	0.59	0.41	0.52	0.60	0.63	0.65	0.57	0.72	0.70	1								
Food supply (kcal/capita/day)	0.64	0.41	0.45	0.53	0.82	0.82	0.61	0.76	0.79	0.62	1							
Childhood stunting (%)	-0.63	-0.40	-0.42	-0.59	-0.74	-0.77	-0.54	-0.78	-0.84	-0.61	-0.71	1						
ln(Childhood stunting) (%)	-0.57	-0.48	-0.40	-0.54	-0.81	-0.79	-0.69	-0.80	-0.85	-0.58	-0.71	0.92	1					
Pure alcohol consumption (L/capita/year)	0.19	0.33	0.09	0.23	0.57	0.53	0.43	0.50	0.53	0.27	0.42	-0.41	-0.50	1				
KOF Index (total)	0.48	0.48	0.34	0.58	0.80	0.77	0.67	0.79	0.84	0.63	0.76	-0.70	-0.75	0.59	1			
KOF Index (economical dimension)	0.51	0.45	0.28	0.53	0.70	0.68	0.64	0.75	0.76	0.53	0.62	-0.58	-0.64	0.52	0.83	1		
KOF Index (social dimension)	0.67	0.61	0.45	0.63	0.83	0.83	0.70	0.91	0.93	0.62	0.72	-0.82	-0.82	0.56	0.83	0.82	1	
KOF Index (political dimension)	0.09	0.16	0.15	0.32	0.45	0.40	0.32	0.32	0.40	0.39	0.52	-0.29	-0.36	0.36	0.73	0.27	0.29	1

Figure A - 2: Pairwise correlation plot all variables cf. Table B - 1



Note: SOM6 presents the number of fast food restaurants per million inhabitants. Other variables are considered self-explanatory in combination with *Section 3.4*.

Table A - 3: VIF scores model 1 and 2

VIF scores	Model-1	Model-2
Human Development Index (HDI)	5.73	3.52
ln(Fat supply) (kg/capita/year)	4.27	2.76
Urban population (% of total population)	2.94	2.69
Sugar consumption (kg/capita/year)	2.17	2.58
Adults (18+) physically inactive (%)	1.60	1.70
KOF Index (economical dimension)	2.62	1.73
KOF Index (political dimension)	1.48	1.15
Mean VIF	2.97	2.30

The Variance Inflation Factor (VIF) confirms the pairwise correlation matrices and plots, collinearity is not a problem. All models comply with the VIF criteria. Acceptable values for VIF are between 1 and 10, with optimum values smaller than 5. The average VIF values would need to be of the order of 3 or less (Bowerman, O'Connell & Murphree, 2011).

Figure A - 3: Check heteroscedasticity model 1 (all countries)

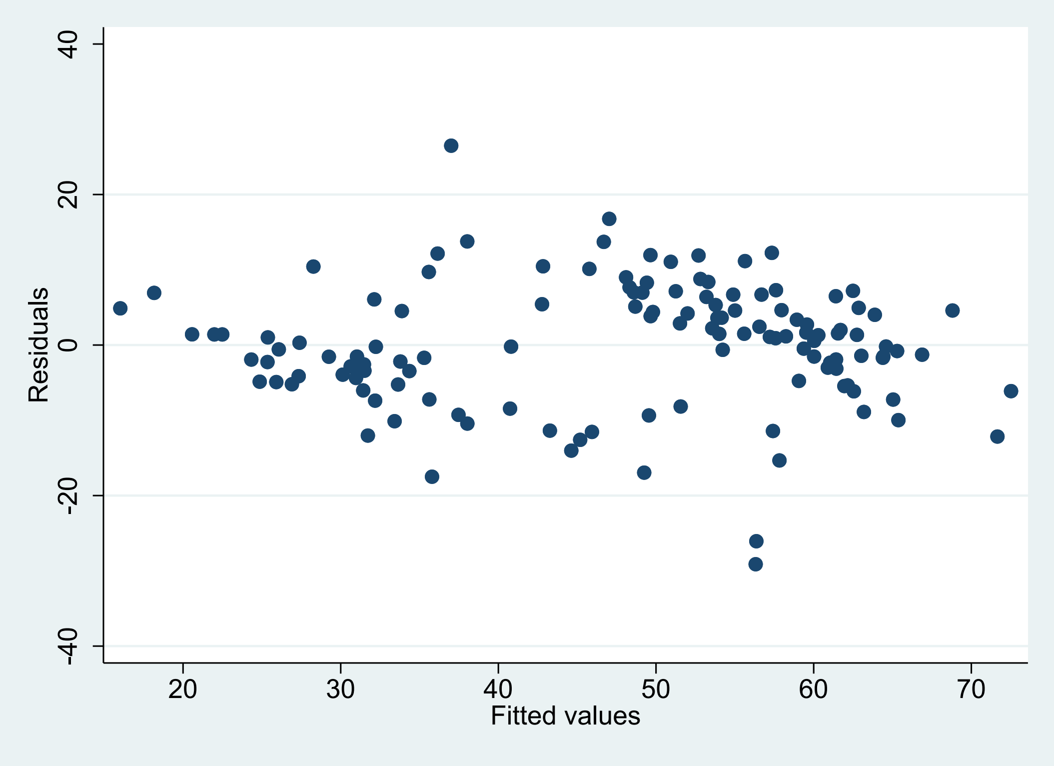


Figure A - 4: Check heteroscedasticity model 2 (developing countries only)

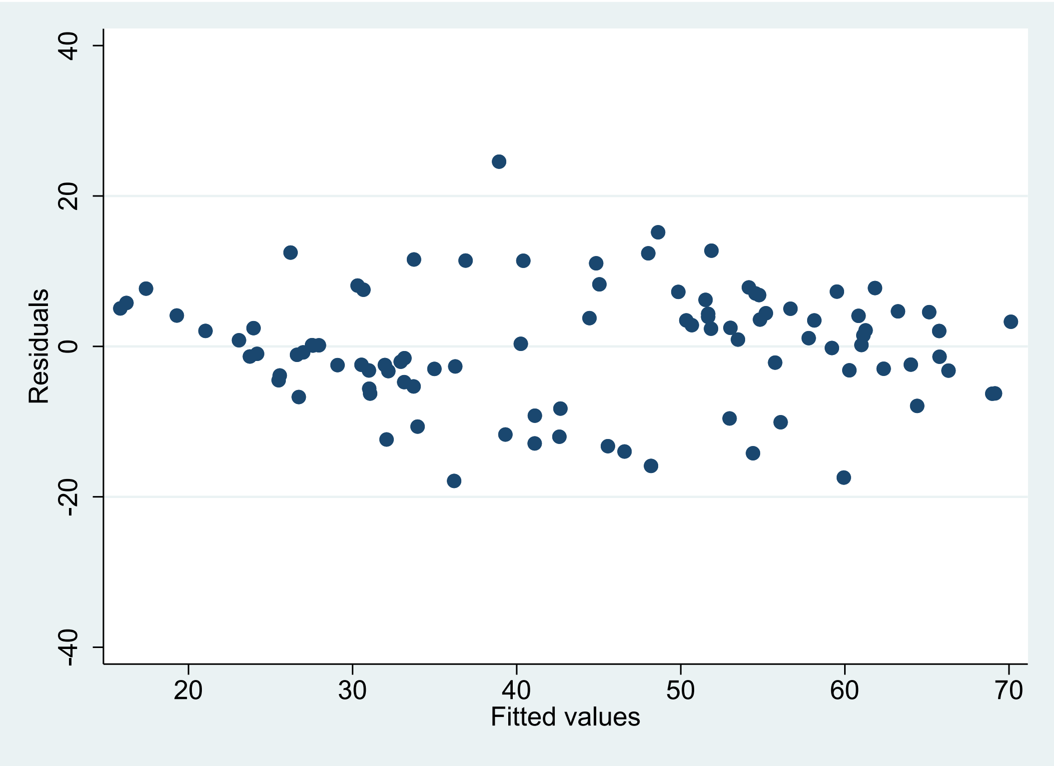


Figure A - 5: Check missing data overweight prevalence

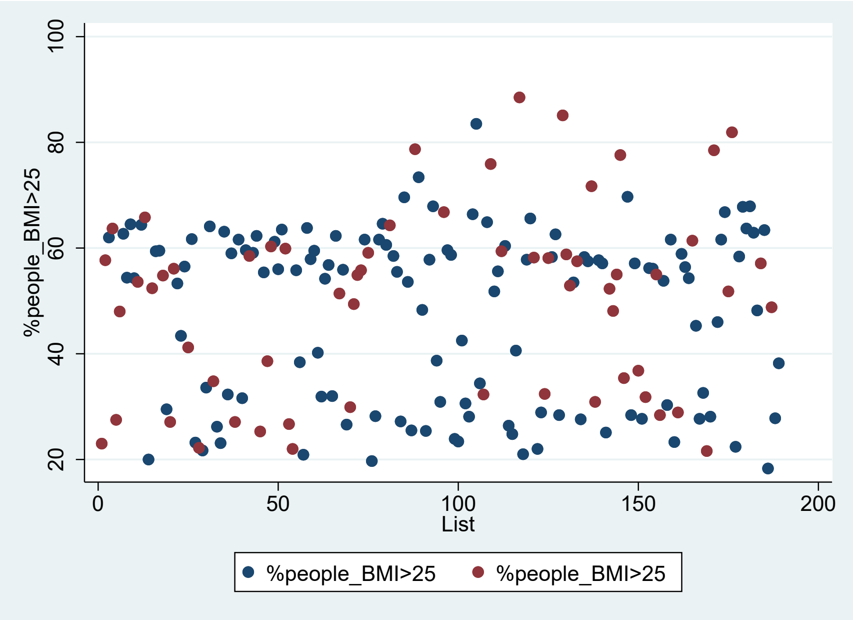


Figure B - 1 shows the distribution of the missing vs included data in the regression analysis for the variable prevalence of overweight (= %people_BMI>25) versus the countries ranked in alphabetical order (1= Afghanistan, 189 = Zimbabwe). No clear patterns distinguished (red = missing, blue = included).

Figure A - 6: Check missing data physical inactivity

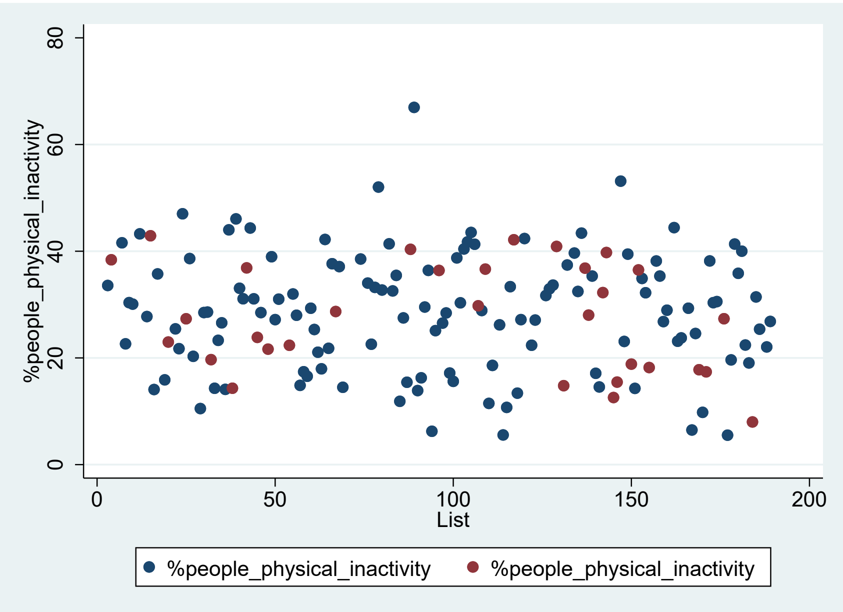


Figure B - 2 shows the distribution of the missing vs included data in the regression analysis for the variable physical inactivity (= %people_physical_inactivity) versus the countries ranked in alphabetical order (1= Afghanistan, 189 = Zimbabwe). No clear patterns distinguished (red = missing, blue = included).

Figure A - 7: Check missing data sugar supply

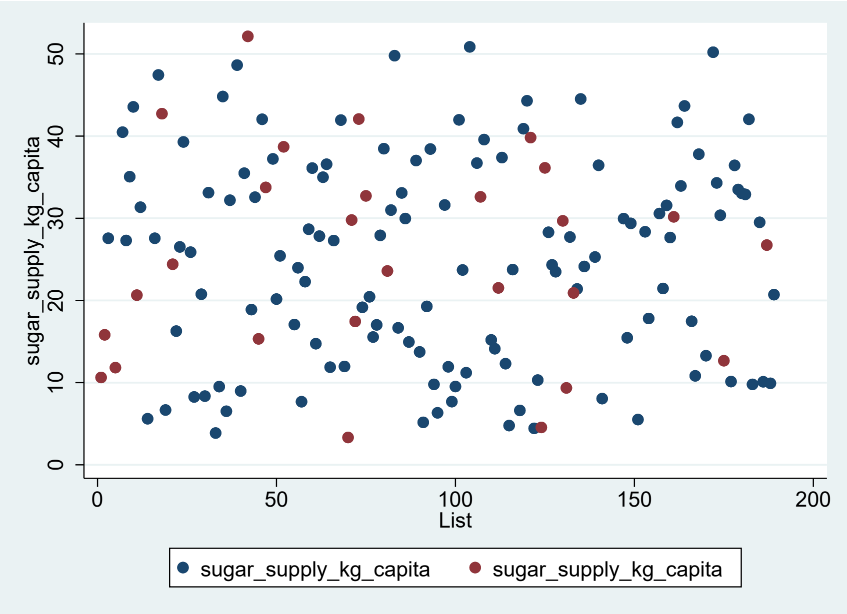


Figure B - 3 shows the distribution of the missing vs included data in the regression analysis for the variable sugar supply (= sugar_supply_kg_capita) versus the countries ranked in alphabetical order (1= Afghanistan, 189 = Zimbabwe). No clear patterns distinguished (red = missing, blue = included).

Figure A - 8: Check missing data fat supply

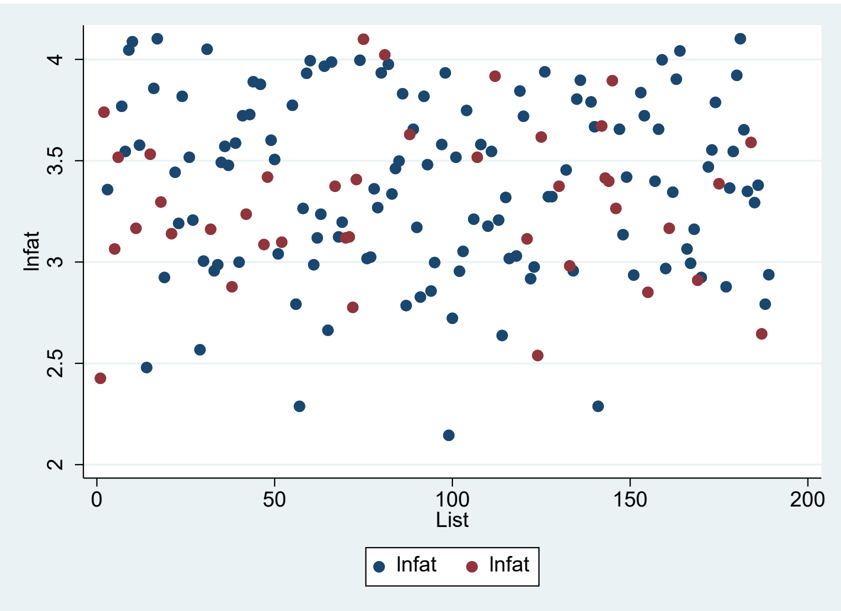


Figure B - 4 shows the distribution of the missing vs included data in the regression analysis for the variable (natural logarithm of) fat supply (= lnfat) versus the countries ranked in alphabetical order (1= Afghanistan, 189 = Zimbabwe). No clear patterns distinguished (red = missing, blue = included).

Appendix B – Background information

Table B - 1: All independent variables identified from literature

Independent variable[1)]	Justification (refer to *Chapter 2*)
1. Total no. fast food restaurants/1 million inhabitants	Davis & Carpenter (2009) Burgoine, Sarkar, Webster & Monsivais (2018)
2. Adults (18+) physically inactive (%)	Hills, Andersen & Byrne (2011) Ladabaum, Mannalithara, Myer & Singh (2014)
3. Sugar available for consumption (kg/capita/year)	Hu, F. (2013) Stanhope (2016) Du, Tugendhaft, Erzse & Hofman (2018)
4. ln(Fat supply) (kg/capita/year)*	Golay & Bobbioni (1997) Mensink, Zock, Kester & Katan (2003) Mozaffarian et al. (2004) Cohut (2018) Hu, S. et al. (2018)
5. ln(GDP per capita) (current USD)*	Kinge, Strand, Vollset and Skirbekk (2015)
6. Human Development Index (HDI)	Ataey, Jafarvand, Adham and Moradi-Asl (2020)
7. Urban population (% of total population)	Popkin (2019) Kuddus, Tynan and McBryde (2020) Tian, Zhao, Li, Wang and Shi (2014) Cooksey-Stowers, Schwartz and Brownell (2014) Ghosh-Dastidar et al. (2014)
8. Food supply (kcal/capita/day)	WHO (2003) Silventoinen et al. (2004) Wright & Aronne (2012)
9. Pure alcohol consumption (L/capita/year)	National Health Service of the United Kingdom (2019) Sayon-Orea, Martinez-Gonzalez & Bes-Rastrollo (2011) Traversy & Chaput (2015)
10. ln(Childhood stunting) (%)*	van der Merwe & Pepper (2006) Muhammad (2018)
11. KOF Index (total lumped sum)	Costa-Font & Mas (2016)
12. KOF Index (total economic)	Costa-Font & Mas (2016)
13. KOF Index (total social)**	Costa-Font & Mas (2016)
14. KOF Index (total political)	Costa-Font & Mas (2016)

Notes:
[1)] All details on the independent variables such as definitions, methods of measurement, sources and year of observation can be found in *Section 3.4*

Table B - 2: Sources of independent variables as indentified from literature

Independent variable[1)]	Year of observation[2)]	Data source[3)]
Prevalence of overweight among adults, BMI ≥ 25 (%)	2016	(WHO, 2021c)
Prevalence of obesity among adults, BMI ≥ 30 (%)	2016	(WHO, 2021d)
Fast food restaurants	2021	Company websites[4)]
Prevalence of insufficient physical activity among adults 18+ years	2016	(WHO, 2021e)
Sugar available for consumption (kg/capita/year)	2016	(FAOSTAT, 2021a)
Fat supply (kg/capita/year)	2016	(FAOSTAT, 2021b)
GDP per capita in current United States Dollars (USD)	2016	(World Bank, 2021a)
Human Development Index (HDI)	2016	(UNDP, 2021)
Urban population (% of total population)	2016	(World Bank, 2021b)
Food supply (kcal/capita/day)	2016	(FAOSTAT, 2021c)
Stunting prevalence children < 5 years (% height for age <-2 SD)[5)]	2016	(WHO, 2021f)
Population (total world)	2016	(World Bank, 2021c)
Pure alcohol consumption (L/capita/year)	2016	(WHO, 2021g)
KOF Globalization Indices	2016	(KOF, 2022)

Notes:

1) The variables are retrieved from databases as referred to in the column "Data source". Prevalence of overweight and obesity are taken from the referred data source for all years considered.
2) The year of observation represents the year for which the data was published in the database.
3) The complete references of the data sources can be found in the References section.
4) Fast food chains included: McDonalds (2021), Subway (2021) and KFC (2021).
5) <-2 SD indicating below 2 standard deviations of the mean.

All details of the parameters identified in literature, but not included in the regression analysis are given below.

<u>Prevalence of obesity</u>

The prevalence of obesity presents the percentage of adults aged 18 years and older of a country's population with a BMI≥30 (age-standardized estimate). BMI is calculated by dividing the persons weight by its height squared. The data are retrieved from the WHO database for the year 2016 (WHO, 2021c). The dataset from the WHO database originates originally from NCD Risk Factor Collaboration (NCD-RisC), which published their results in The Lancet of October 2017 (NCD Risk Factor Collaboration, 2017). The prevalence of obesity is obtained from 2416 population-based studies with measurements of height and weight on 97.4 million participants aged 20 years and older.

<u>Food supply</u>

Food supply presents the total national food available for consumption. Food supply data are taken from FAOSTAT for the year 2016 (FAOSTAT, 2021c). Food supply is not a direct measurement of the actual food consumption, but calculated from the quantity on stocks + quantities imported and produced – quantities exported, seed, fed to animals, waste and other non-food uses – ending stocks. The amount is expressed in kcal/capita/day for each country included in the database.

The food balance sheets, wherefrom the data are obtained, are derived statistics, of which the reliability depends of the underlying basic statistics of population, supply and utilization of foods and/or of their nutritive value. Often the data are supported by country's own expert opinions. Quality and coverage vary across countries and commodities

GDP per capita

Gross domestig product (GDP) per capita is defined as the gross domestic product divided by midyear population. It is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. Depreciation of fabricated assets or for depletion and degradation of natural resources is not included. The data are retrieved from the World Bank database for the year 2016 (World Bank, 2021a). The GDP is expressed in current U.S. dollars.

Pure alcohol consumption

Total alcohol consumption per capita (APC) is defined as the total sum of consumed alcohol per adult aged 15 years and older over one calender year. The data are retrieved from the WHO database (WHO, 2021g). It is calculated as the three-year average of recorded and unrecorded APC, adjusted for three-year average tourist consumption. Recorded alcohol consumption is based on official statistics (production, import, export, and sales or taxation data), while the unrecorded alcohol consumption is estimated. Tourist consumption takes into account tourists visiting the country and inhabitants visiting other countries and is based on United Nations tourist statistics. The total alcohol consumption is expressed as the total liters pure alcohol consumed per capita in one calender year.

Childhood stunting

Childhood stunting refers to children aged 5 years or less who are too short for their age as a result of chronic or recurrent malnutrition. Stunting is defined as the percentage of children aged 5 years or less who's height is 2 standard deviations below the median of the WHO Child Growth Standards. The data are retrieved from the WHO database for the year 2016 (WHO, 2021f). Survey estimates are based on standardized methodology according to the WHO Child Growth Standards and referred to as child malnutrition country servey results (JME). If a country has no data available it is not included in the database.

Table B - 3: Descriptive statistics all variables from literature

Variable	Observations	Mean	Std. Dev.	Min	Max
Population per country in 2001	189	32526369	1.247e+08	9478	1.272e+09
Population per country in 2006	189	34623816	1.314e+08	9827	1.311e+09
Population per country in 2011	189	36777771	1.376e+08	10069	1.344e+09
Population per country in 2016	189	38998337	1.435e+08	10474	1.379e+09
No. Subway restaurants per country	189	195.2	1572.0	0.0	21246
No. McDonalds restaurants per country	189	191.3	1081.7	0.0	14146
No. KFC restaurants per country	189	119.1	553.2	0.0	6078
Total no. fast food restaurants per country	189	505.7	3001.9	0.0	39454
Total no. fast food restaurants per 1 million inhabitants	189	14.0	23.4	0.0	139.2
Adults (18+) physically inactive (%)	159	28.4	10.9	5.5	67.0
Adults (18+) with BMI≥25 2001 (%)	189	40.3	17.0	10.5	85.4
Adults (18+) with BMI≥25 2006 (%)	189	43.0	17.1	12.7	86.5
Adults (18+) with BMI≥25 2011 (%)	189	45.8	17.0	15.3	87.6
Adults (18+) with BMI≥25 2016 (%)	189	48.6	16.8	18.3	88.5
Adults (18+) with BMI≥30 2001 (%)	189	13.6	9.3	0.7	55.5
Adults (18+) with BMI≥30 2006 (%)	189	15.4	9.9	1.0	57.4
Adults (18+) with BMI≥30 2011 (%)	189	17.4	10.4	1.5	59.3
Adults (18+) with BMI≥30 2016 (%)	189	19.6	10.9	2.1	61
Sugar available for consumption (kg/capita/year)	154	25.1	12.5	3.3	52.1
Fat supply (kg/capita/year)	167	31.5	12.8	8.5	60.5
ln(Fat supply)	167	3.4	0.4	2.1	4.1
GDP per capita (current USD)	185	12675.0	17533.9	282.2	104278.4
ln(GDP per capita)	185	8.6	1.4	5.6	11.6
Human Development Index (HDI)	184	0.8	0.2	0.4	1.0
Urban population (% of total population)	188	58.1	22.7	12.4	100
Food supply (kcal/capita/day)	167	2848.5	457.2	1745	3774
Pure alcohol consumption (L/capita/year)	185	5.8	4.1	0.0	14.3
Childhood stunting (%)	149	19.8	13.4	1.3	56.3
ln(Childhood stunting)	149	2.7	0.9	0.3	4.0
KOF Globalization Index	186	61.9	14.5	30.3	90.7
KOF Economic Globalization Index	183	57.5	16.5	26.6	93.9
KOF Social Globalization Index	186	63.2	17.1	26.9	91.8
KOF Political Globalization Index	186	65.6	21.1	11.0	97.9

***Table B - 4:* Regression tables model 1 with and without KOF economical dimension**

Dependent variable: Overweight Prevalence Adults (18+) BMI≥25 (%)	Model-1 Complete Case		Model-1 Complete Case	
Adults (18+) physically inactive (%)	0.038	*0.656*	0.054	*0.523*
	(0.086)		(0.084)	
Sugar consumption (kg/capita/year)	0.286	*0.001******	0.272	*0.002******
	(0.088)		(0.084)	
ln(Fat supply) (kg/capita/year)	9.685	*0.007******	9.272	*0.008******
	(3.542)		(3.454)	
Human Development Index (HDI)	22.462	*0.053***	17.415	*0.083***
	(11.487)		(9.962)	
Urban population (% of total population)	0.268	*0.000******	0.275	*0.000******
	(0.058)		(0.057)	
KOF Index (economical dimension)	-0.063	*0.371*		
	(0.070)			
KOF Index (political dimension)	-0.137	*0.017*****	-0.133	*0.020*****
	(0.057)		(0.056)	
Constant	-12.128	*0.095***	-11.543	*0.104*
	(7.210)		(7.042)	
Number of variables K and ratio N/K	7 [18.3]		6 [21.5]	
Number of observations N	128		129	
R^2	0.741		0.740	
Countries included	Developed and developing		Developed and developing	

Standard errors in parentheses. *P-values in italics, * p < 0.10, ** p < 0.05, *** p < 0.01*

Table B - 5: Regression tables model 2 with and without KOF economical dimension

Dependent variable: Overweight Prevalence Adults (18+) BMI≥25 (%)	Model-2		Model-2	
Adults (18+) physically inactive (%)	-0.011 (0.091)	*0.908*	-0.004 (0.089)	*0.967*
Sugar consumption (kg/capita/year)	0.205 (0.106)	*0.056***	0.188 (0.098)	*0.059***
ln(Fat supply) (kg/capita/year)	7.211 (3.723)	*0.056***	7.336 (3.651)	*0.048****
Human Development Index (HDI)	33.355 (11.664)	*0.005*****	32.061 (10.630)	*0.003*****
Urban population (% of total population)	0.349 (0.063)	*0.000*****	0.350 (0.062)	*0.000*****
KOF Index (economical dimension)	-0.024 (0.078)	*0.756*		
KOF Index (political dimension)	-0.078 (0.058)	*0.184*	-0.075 (0.057)	*0.191*
Constant	-17.970 (9.143)	*0.053***	-18.827 (8.858)	*0.036****
Number of variables K and ratio N/K	7 [13.7]		6 [16.2]	
Number of observations N	96		97	
R^2	0.782		0.784	
Countries included	Developing		Developing	

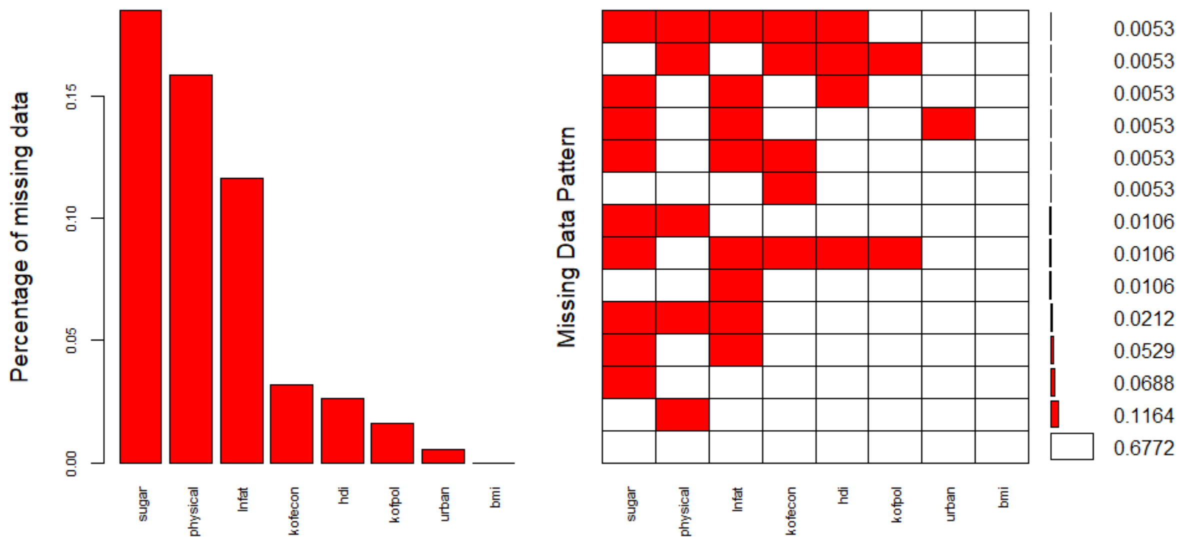
Standard errors in parentheses. *P-values in italics, * p < 0.10, ** p < 0.05, *** p < 0.01*

Appendix C – Data imputation

Data imputation is carried out using R Studio.

Mostly the countries that are excluded do not miss data on all factors. Mostly Information lacks for one or two of the factors only (*Figure C - 1*). The total set of countries of the world is 194. The set included in the analysis yields 189 countries, because Liechtenstein, Monaco, Palestine State, San Marion and South Sudan are removed in view of lacking data in BMI. From this set 61 countries lack data for one or more variables (*Figure C - 1*).

Figure C - 1: Statistics of missing data all countries included



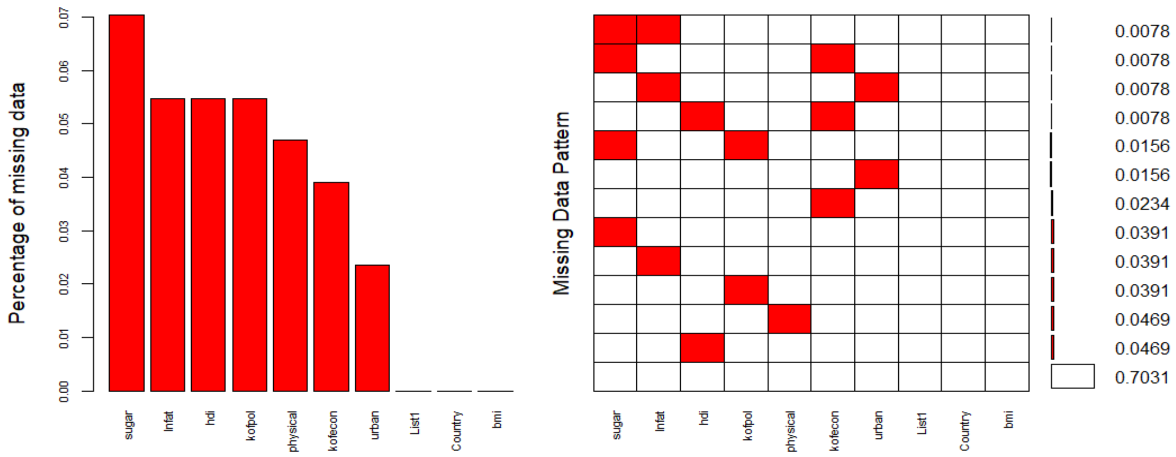
From *Figure C - 1*, it can be observed that the fraction of countries with complete data, white color in the right picture, is 0.6772 (68%). This means that about 32% of the data is excluded from the analysis if a complete case analysis is carried out. From *Figure C - 1*, it can also be seen that the total of countries missing 4 or more out of the 7 factors yields only 2% of the total number if countries. In other words, the majority of the countries with missing data on factor (93%) contain still information of 4 or more of the factors. Excluding countries for the reason of missing data on one or more of the factors, the complete case analysis, implies that statistical interesting information is left out, while actually still useful information is present. It is decided to impute the missing data. The missing data does not seem to show a clear pattern, other than that sugar, physical and lnfat are the factors that miss the most data. But in the missingness no clear patterns was discovered (*Figure A - 6*, *Figure A - 7*, *Figure A - 8*). It is considered likely that imputation can result in reliable results when making use of information already hidden in the complete data set. The countries with one or more missing data, unlikely behave different than the other countries given that they miss data.

To confirm this assumption, the impact and effectiveness of imputation is investigated for the dataset under consideration using different imputation methods.

This is done by taking the 128 countries with full information using as the reference. This is a complete set, with true data as originally reported in the source databases. This data set is referred to as the original set.

Data are removed from this original set, at random, to obtain a new base set. This base set still contains 128 countries, but 30% of the countries miss one or more of the factors include due the ad random removal of data (*Figure C - 2*).

Figure C - 2: Base set created by removing 30% from the original set



In the next step of the check of the imputation process, the base set is imputed 100 times to generate 100 different of complete sets. The imputes set are analyzed and pooled to produce a final pooled regression result.

It the multiple imputation is successful, than it would logic thinking to assume that this imputed set will produce more or less the same test results as the original, as it is derived from this original set by artificially removing data. The advantage of this procedure is that the imputed base set can be compared directly with the original set. If the procedure is successful, than the process can also produce reliable results when applied to the 189 countries, provided that the data are not missing for a specific reason that also affect the research results. As was discussed before, this is considered unlikely to be the case.

First a regression is carried out on the original data set, comprising of the 128 countries with data available for all variables, in both Stata and R to compare the regression outcomes. The results were identical.

The set that includes 128 countries with complete data available is the original set, the base set includes 90 countries after artificial removal (at random) of 38 countries. The imputed set results from the imputation to a complete set and thus includes 128 countries again.

Table C - 1: Check PMM imputation method

Dependent variable: Overweight Prevalence Adults (18+) BMI≥25 (%)	Original Set		Base Set		Imputed Base Set PMM[1)]	
Adults (18+) physically inactive (%)	0.038 (0.086)	*0.656*	-0.019 (0.115)	*0.870*	0.014 (0.088)	*0.870*
Sugar consumption (kg/capita/year)	0.286 (0.088)	*0.001******	0.365 (0.118)	*0.003******	0.324 (0.089)	*0.000******
ln(Fat supply) (kg/capita/year)	9.685 (3.542)	*0.007******	9.449 (4.429)	*0.036****	9.118 (3.560)	*0.012****
Human Development Index (HDI)	22.462 (11.487)	*0.053**	22.964 (16.397)	*0.165*	26.534 (11.634)	*0.024****
Urban population (% of total population)	0.268 (0.058)	*0.000******	0.272 (0.008)	*0.001******	0.254 (0.059)	*0.000******
KOF Index (economical dimension)	-0.063 (0.070)	*0.371*	-0.116 (0.093)	*0.213*	-0.097 (0.072)	*0.183*
KOF Index (political dimension)	-0.137 (0.057)	*0.017****	-0.136 (0.076)	*0.077**	-0.138 (0.056)	*0.016****
Constant	-12.128 (7.210)	*0.095**	-9.654 (9.537)	*0.314*	-10.690 (7.215)	*0.141*
Number of variables K and ratio N/K	7 [18.3]		7 [12.9]		7 [18.3]	
Number of observations N	128		90		90 + 38 imputed	
R^2	0.741		0.699		0.745	
Countries included	Developed and developing		Developed and developing		Developed and developing	

Standard errors in parentheses. *P-values in italics, * p < 0.10, ** p < 0.05, *** p < 0.01*

Notes:
1) PMM = Predicted Mean Matching (100 sets generated, 25 iterations, 5 donor sets, all predictors included)

From *Figure C - 3*, it can be concluded that the imputed sets cover the original density plot quite well (although not visible anymore as it is hidden behind the imputed sets). From *Table C - 2*, it can be concluded that both, PMM and Bayesian imputation perform well if applied to the set with randomly removed data. Although the differences between PMM and Bayesian regression appeared to be very small, the literature recommends PPM as an all-around method with relatively small biased estimates providing a better performance than many other methods (van Buuren, 2018).

Figure C - 3: Density plots of all 100 imputed data sets

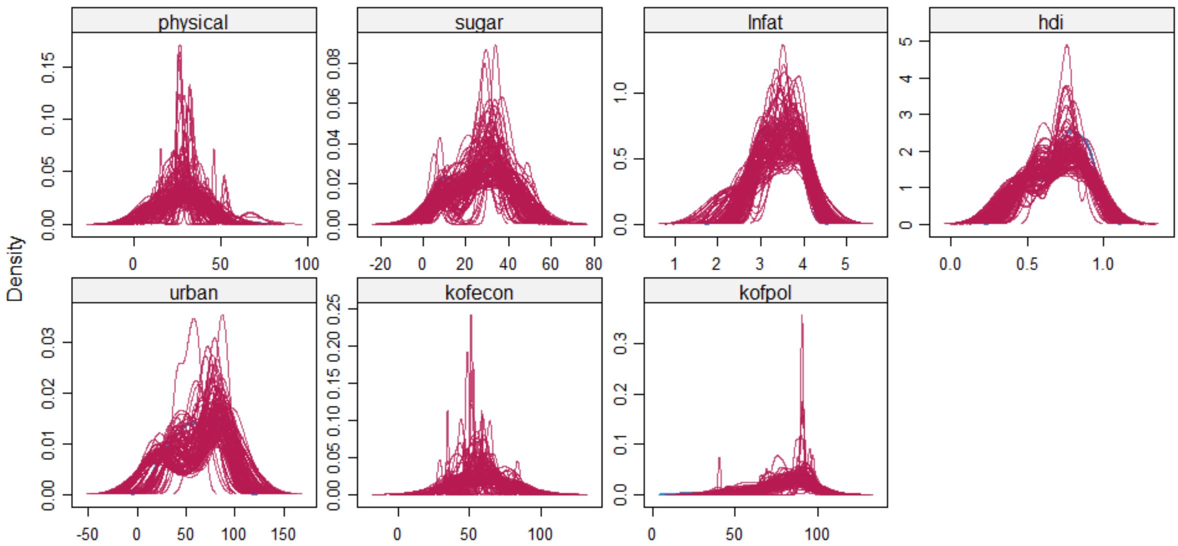


Table C - 2: Performance check PMM and Bayesian imputation

Dependent variable: Overweight Prevalence Adults (18+) BMI≥25 (%)	Imputed Base Set PMM[1)]		Imputed Base Set Bayesian regression[2)]		Original Set	
Adults (18+) physically inactive (%)	0.016	*0.852*	0.015	*0.861*	0.038	*0.656*
	(0.088)		(0.088)		(0.086)	
Sugar consumption (kg/capita/year)	0.318	*0.001****	0.316	*0.001****	0.286	*0.001****
	(0.089)		(0.090)		(0.088)	
ln(Fat supply) (kg/capita/year)	9.133	*0.012***	9.089	*0.013***	9.685	*0.007****
	(3.568)		(3.596)		(3.542)	
Human Development Index (HDI)	26.409	*0.026***	26.267	*0.028***	22.462	*0.053**
	(11.681)		(11.815)		(11.487)	
Urban population (% of total population)	0.255	*0.000****	0.255	*0.000****	0.268	*0.000****
	(0.060)		(0.060)		(0.058)	
KOF Index (economical dimension)	-0.093	*0.202*	-0.090	*0.220*	-0.063	*0.371*
	(0.073)		(0.073)		(0.070)	
KOF Index (political dimension)	-0.136	*0.017***	-0.134	*0.019***	-0.137	*0.017***
	(0.056)		(0.056)		(0.057)	
Constant	-10.868	*0.134*	-10.990	*0.136*	-12.128	*0.095**
	(7.206)		(7.327)		(7.210)	
Number of variables K and ratio N/K	7 [18.3]		7 [18.3]		7 [18.3]	
Number of observations N	90 + 38 imputed		90 + 38 imputed		90 + 38 imputed	
R^2	0.744		0.742		0.742	
Countries included	Developed and developing		Developed and developing		Developed and developing	

Standard errors in parentheses. *P-values in italics, * p < 0.10, ** p < 0.05, *** p < 0.01*

Notes:
[1)] PMM (Predicted Mean Matching) = 100 sets generated, 25 iterations, 5 donor sets, all predictors included
[2)] Bayesian regression = 100 sets generated, 25 iterations, all predictors included

Appendix D – KOF structure

Figure D - 1: KOF structure

2021 Globalisation Index: Structure, variables and weights

Globalisation Index, de facto	Weights	Globalisation Index, de jure	Weights
Economic Globalisation, de facto	***33.3***	***Economic Globalisation, de jure***	***33.3***
Trade Globalisation, de facto	*50.0*	*Trade Globalisation, de jure*	*50.0*
Trade in goods	37.2	Trade regulations	26.8
Trade in services	43.0	Trade taxes	28.1
Trade partner diversity	19.8	Tariffs	27.1
		Trade agreements	18.0
Financial Globalisation, de facto	*50.0*	*Financial Globalisation, de jure*	*50.0*
Foreign direct investment	26.3	Investment restrictions	30.2
Portfolio investment	16.7	Capital account openness	39.0
International debt	28.6	International Investment Agreements	30.8
International reserves	1.0		
International income payments	27.4		
Social Globalisation, de facto	***33.3***	***Social Globalisation, de jure***	***33.3***
Interpersonal Globalisation, de facto	*33.3*	*Interpersonal Globalisation, de jure*	*33.3*
International voice traffic	20.7	Telephone subscriptions	39.1
Transfers	22.1	Freedom to visit	32.4
International tourism	21.1	International airports	28.6
International students	19.0		
Migration	17.2		
Informational Globalisation, de facto	*33.3*	*Informational Globalisation, de jure*	*33.3*
Used internet bandwidth	40.7	Television access	37.7
International patents	29.6	Internet access	43.3
High technology exports	29.6	Press freedom	19.0
Cultural Globalisation, de facto	*33.3*	*Cultural Globalisation, de jure*	*33.3*
Trade in cultural goods	28.6	Gender parity	22.5
Trade in personal services	24.8	Human capital	41.7
International trademarks	7.9	Civil liberties	35.8
McDonald's restaurant	22.0		
IKEA stores	16.8		
Political Globalisation, de facto	***33.3***	***Political Globalisation, de jure***	***33.3***
Embassies	37.2	International organisations	36.5
UN peace keeping missions	24.7	International treaties	32.6
International NGOs	38.2	Treaty partner diversity	30.9

Notes: Weights in percent for the year 2019. Weights for the individual variables are time variant.
Overall indices for each aggregation level are calculated by the average of the respective de facto and de jure indices.

Source: https://ethz.ch/content/dam/ethz/special-interest/dual/kof-dam/documents/Globalization/2021/KOFGI_2021_structure_19.pdf

Appendix E – Sensitivity analysis KOF indices

Table E - 1: Sensitivity analysis KOF indices (with and without HDI)

Dependent variable: Overweight Prevalence Adults (18+) BMI≥25 (%)	Model-1 Complete Case+)		Model-1 (Sensitivity 1) Complete Case++) KOF social added		Model-1 (Sensitivity 2) Complete Case+++) KOF social replacing HDI	
Adults (18+) physically inactive (%)	0.038 (0.086)	*0.656*	0.040 (0.086)	*0.644*	0.046 (0.087)	*0.600*
Sugar consumption (kg/capita/year)	0.286 (0.088)	*0.001****	0.293 (0.089)	*0.001****	0.289 (0.090)	*0.002****
ln(Fat supply) (kg/capita/year)	9.685 (3.542)	*0.007****	10.076 (3.683)	*0.007****	10.614 (3.696)	*0.005****
Human Development Index (HDI)	22.462 (11.487)	*0.053**	27.122 (16.264)	*0.098**		
Urban population (% of total population)	0.268 (0.058)	*0.000****	0.269 (0.058)	*0.000****	0.287 (0.058)	*0.000****
KOF Index (social dimension)			-0.073 (0.180)	*0.685*	0.139 (0.129)	*0.283*
KOF Index (economical dimension)	-0.063 (0.070)	*0.371*	-0.042 (0.087)	*0.632*	-0.062 (0.087)	*0.478*
KOF Index (political dimension)	-0.137 (0.057)	*0.017***	-0.141 (0.057)	*0.016***	-0.121 (0.057)	*0.035***
Constant	-12.128 (7.210)	*0.095**	-13.322 (7.809)	*0.091**	-10.614 (7.695)	*0.170*
Number of variables K and ratio N/K	7 [18.3]		8 [16.0]		7 [18.3]	
Number of observations N	128		128		128	
R^2	0.741		0.742		0.735	
Adj. R^2	0.726		0.724		0.720	
Countries included	Developed and developing		Developed and developing		Developed and developing	

Standard errors in parentheses. *P-values in italics, * p < 0.10, ** p < 0.05, *** p < 0.01*
+) The F-test complete case produced the following results: F(7, 120) = 49.07, p-value = 0.000
++) The F-test complete case produced the following results: F(8, 119) = 42.66, p-value = 0.000
+++) The F-test complete case produced the following results: F(7, 120) = 47.65, p-value = 0.000

It can be observed that adding the KOF index social dimension to the complete case mainly affects the significance for the KOF index economical dimension and HDI, suggesting that KOF index social and HDI are interchangeable. It is noted that the pairwise correlation was 0.93 (one of the reasons why KOF social was left out in the first place, refer to *Section 3.4*).

HDI and KOF social dimension are to some extent interchangeable. The impact of this change on the other dominant variables, sugar supply, fat supply (if included) and urbanization is usually small. This can be observed by studying the results from the sensitive analyses produced in *Table E - 1* and *Table E - 2*. The impact on other factors that matter remains relatively small for the combinations investigated. It is noted that the KOF social dimension is only significant if ln(fat supply) is excluded. But it is noted that ln(fat supply) is highly correlated with both KOF social dimension and HDI (0.83 in both cases).

Table E - 2: Sensitivity analysis KOF indices

Dependent variable: Overweight Prevalence Adults (18+) BMI≥25 (%)	Model-1 Complete Case+)		Model-1 (Sensitivity 3) Complete Case++)		Model-1 (Sensitivity 4) Complete Case+++)	
Adults (18+) physically inactive (%)	0.038 (0.086)	*0.656*	0.056 (0.091)	*0.543*	0.065 (0.093)	*0.489*
Sugar consumption (kg/capita/year)	0.286 (0.088)	*0.001******	0.286 (0.093)	*0.003******	0.281 (0.097)	*0.004******
ln(Fat supply) (kg/capita/year)	9.685 (3.542)	*0.007******				
Human Development Index (HDI)	22.462 (11.487)	*0.053**	36.963 (11.163)	*0.001******		
Urban population (% of total population)	0.268 (0.058)	*0.000******	0.284 (0.059)	*0.000******	0.311 (0.059)	*0.000******
KOF Index (social dimension)					0.301 (0.121)	*0.014****
KOF Index (economical dimension)	-0.063 (0.070)	*0.371*	-0.020 (0.074)	*0.784*	-0.048 (0.093)	*0.606*
KOF Index (political dimension)	-0.137 (0.057)	*0.017****	-0.110 (0.058)	*0.060**	-0.080 (0.058)	*0.167*
Constant	-12.128 (7.210)	*0.095**	4.375 (4.649)	*0.348*	9.497 (4.357)	*0.031****
Number of variables K and ratio N/K	7 [18.3]		6 [21.7]		6 [21.7]	
Number of observations N	128		130		130	
R^2	0.741		0.703		0.692	
Adj. R^2	0.726		0.689		0.677	
Countries included	Developed and developing		Developed and developing		Developed and developing	

Standard errors in parentheses. *P-values in italics, * p < 0.10, ** p < 0.05, *** p < 0.01*
+) The F-test complete case produced the following results: F(7, 120) = 49.07, p-value = 0.000
++) The F-test complete case produced the following results: F(6, 123) = 48.57, p-value = 0.000
Original but, without ln(fat supply)
+++) The F-test complete case produced the following results: F(6, 123) = 46.12, p-value = 0.000
Original but, without ln(fat supply) and HDI replaced by KOF Index (social dimension)