Long-Term Effects of Extreme Events on Nutrient Intake:

Evidence from The Great Famine in China

Milorad Plavšić

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

School of Agriculture, Policy and Development

Department of Applied Economics and Marketing

28th June 2022

Declaration of Original Authorship

I declare that this research is my own original work and all citations from other sources have been acknowledged.

Date: 28/06/2022

Signed: Milorad Plavšić

Abstract

Famines claimed more than 70 million lives in the 20th century (Devereux, 2000). While many efforts were made in the past several decades to prevent future famines and other forms of food shortages, global trends, such as pandemics, more frequent extreme weather patterns, and armed conflicts suggest that famines are not a matter of the past but will remain a looming risk in the future.

Short-term and long-term effects of famines have been extensively studied. One of those short-term effects is the excess mortality rate caused by famine, and it includes those who died due to starvation and those who were not born due to famine and who would have been born in the absence of famine. Long-term effects of a famine covered by the literature include economic and health effects. Among the economic effects are lost productivity due to lost schooling and reduced cognitive abilities, while the health literature documents long-term impacts on noncommunicable diseases, body mass index (BMI), height, and other indicators of well-being.

Using the Great Famine of China (1959–1961) as a quasi-natural experiment, this dissertation examines the effects of early-life famine on nutrient intake later in life. The Great Famine provides a useful natural laboratory to analyse these effects in that it was very long, with unprecedented severity and substantial variation across the regions of China. Regional variation in famine severity, combined with variations in health consequences across different birth-year cohorts, gender and nationality, provides a suitable setting for identifying the effects of famine on food choice and consumption later in life.

In this dissertation I show that memories of famine have an enduring effect on eating patterns and lifestyle habits. I find evidence that individuals who were more affected by early-life famine have significantly different eating patterns, measured by qualitative and quantitative characteristics. I show that the Great Famine severely affected the dietary diversity and macro-nutrient composition of diets of survivors, in particular, those who were in early childhood during the famine. I also find significant differences across gender, nationality, educational attainment and income, which suggests that some of these factors moderate the harms of famine. Furthermore, I find that exposure to the Great Famine negatively affected food expenditures of famine survivors. Taken together, this study's evidence indicates that the Great Famine had a considerable sustained impact on nutrient intake of the survivors more than 40 years later.

This study contributes to understanding of the long-term effects of early-life adversity on later health and economic outcomes. The policy dimension of this dissertation suggests that, in addition to the emergency food relief, it is necessary to provide famine survivors with long-term support, as their "famine scar" is a very important determinant of their behaviour, leading to a poorer diet throughout their lives.

Acknowledgements

I would first like to thank my supervisors Professor Chittur Srinivasan, and Professor Sanzidur Rahman for their support, guidance and help. They helped me to frame my dissertation, but more importantly they taught me how to think critically, how to be accurate and how to look at the same issue from different angles. I am confident that I will use those skills in my post-PhD era.

I am also grateful to Professor Kelvin Balcombe, Professor Giacomo Zanello, and Professor Ariane Kehlbacher, whose feedback from Confirmation of Registration meeting massively helped me in my research activities and in setting realistic objectives. Additionally, my examiners Professor Balcombe and Professor Calum Turvey provided me with a very specific and constructive input which helped my dissertation get its final shape, and I am grateful to them.

My fellow students from the program Yunbi Mo, Oana Tanasache and Toju Begho and other students from office 304, certainly made my PhD journey more pleasurable. I loved discussing with you research dilemmas, but also non-PhD related topics.

During my PhD journey, I was lucky to work with many incredible people. While working at UKRI, I had a chance to learn from two extraordinary professionals – Professor Nicola Lowe and Dr Tahrat Shahid. Our activities and discussions certainly shaped my dissertation and introduced new dimensions to it. I am currently privileged to work with one of the greatest scholars in areas of food systems and agricultural economics – Professor Prabhu Pingali. I would like to thank Prabhu for bringing new ideas to my work and for supporting me during the final part of my PhD journey. I would also like to express gratitude to the whole TCI team.

My parents, my brother and his family played a very prominent role during my PhD program. I would like to thank them for inspiring me to pursue this program and for providing me the support throughout the journey.

I would like to thank Indjic-Ast family for being there for me and for cheering me on throughout my PhD journey.

Special thanks go to Professor Elizabeth Robinson whose advice were extremely helpful not only while I was working on my research, but extended to my life beyond PhD. Professor Robinson encouraged me to apply to the PhD program, and at times it feels like she believed in me more than I believed myself. I could not have asked for a better PhD advisor.

Finally, my dissertation would not see the light of day if it were not for my Team Umizoomi, who stood strong behind me throughout my PhD journey. My children Dunja and Filip are infinite source of joy, energy, and fun. I publicly pledge to make up for all fun activities, which I missed due to my work on the dissertation. I am indebted for life to my wife Dragana whose love, patience, compassion and encouragement made this dissertation possible. I cannot imagine that I would reach the final stop of my PhD journey without you and your faith in me. Thank you, and I kindly ask you not (ab)use the part "indebted for life" on everyday basis.

Abbreviations

ACE - Adverse Childhood Experiences

ADL - Activities of Daily Living

BMI - Body Mass Index

CCDCP - Chinese Center for Disease Control and Prevention

CFCT - China Food Composition Tables

CFCT – China Food Composition Tables

CHARLS – China Healthy Ageing and Retirement Longitudinal Survey

CHNS – China Health and Nutrition Survey

CSSI - Cohort Size Shrinkage Indices

DDS - Dietary Diversity Score

DQQ - Diet quality questionnaire

EDR - Excess Death Rate

ELSA - English Longitudinal Study of Aging

EUFIC – The European Food Information Council

FAFH – Food Away from Home

FANTA - Food and Technical Assistance Project

FAO – Food and Agriculture Organization of the United Nations

FBS – Food Balance Sheets

FCT – Food Composition Tables

GDP - Gross Domestic Product

HRS - Health and Retirement Survey

IPCC - The Intergovernmental Panel on Climate Change

IV - Instrumental Variables

MetS - Metabolic Syndrome

NBS - China National Bureau of Statistics

NCDs – Non-communicable Diseases

PoCS – Proportion of Coarse Staple Consumption

SDS - Staple Diversity Score

SHARE - Survey of Health, Aging and Retirement in Europe

UN OCHA – The United Nations Office for the Coordination of Humanitarian Affairs

UNCH-CPC – Carolina Population Center at the University of North Carolina at Chapel Hill

WFP - World Food Programme

WHO - World Health Organization

List of figures

List of tables

Contents

1 Background

A recently published report by The Intergovernmental Panel on Climate Change (IPCC) suggests that negative effects of increasing frequency of climate extremes on food production will cause migrations, conflicts and food price spikes (IPCC, 2019). The same report analyses the relationship between land degradation and food insecurity by presenting the case of the Tigray region of Ethiopia, where recurrent drought and famine have occurred. The report notes: “Land degradation and climate change act as threat multipliers for already precarious livelihoods, leaving them [people in degraded areas who directly depend on natural resources for subsistence, food security and income] highly sensitive to extreme climatic events, with consequences such as poverty and food insecurity” (IPCC, 2019). Similarly, in Angola, driven by five consecutive years of drought, in March 2022, nearly 1.6 million people experienced acute levels of food insecurity, which was nearly 60% of the analysed population (UN OCHA, 2022).

Food insecurity can be caused by prolonged factors, as shown above, or temporary factors. Swarms of locusts, a temporary factor, mainly affected cereal production in East Africa and South Asia from 2019 to mid-2022. In contrast, the COVID-19 pandemic, though temporary, affected literally all world regions. While the pandemic started in 2019, it was officially declared in 2020, when most of countries in the world imposed public health measures to reduce its spread. Those measures varied from masking recommendations to complete lockdowns, and it has been recognized that at least some of them might have caused economic and developmental slowdowns (Bauer et al., 2020). Economically advanced and food secure countries had the ability to concentrate their efforts around COVID concerns, while less developed countries tried to balance their needs for COVID public health measures and food security. Ghana, for example, eased already light anti-Covid measures based on food security and poverty concerns, despite a very high number of covid cases (Birner et al., 2021). In 2022, when it seemed that the negative economic and health consequences of the pandemic were contained, China discovered a new strain of virus, and imposed lockdown in Shanghai, a city with the population of 25 million, which sparked serious concerns around food security (Liu, 2022).

Additionally, food shortages can be induced by conflicts and other forms of political turmoil. The main three conflicts that marked 2021 and 2022 are those in Yemen, Afghanistan, and Ukraine. Yemen’s civil war started in 2014, and it is projected that two-thirds of the total population will be food insecure by December 2022, while currently more than 2 million children under 5 and more than 1 million pregnant or breastfeeding women require treatment from acute malnutrition (World Food Programme, 2022a). After two decades of relative stability in Afghanistan, political instability and armed conflict sparked substantial concerns about food insecurity in 2021. It has been estimated that more than half of the Afghani population faces acute food insecurity and almost 9 million people face an emergency level of food insecurity (World Food Programme, 2022b). Unlike the prior two examples of countries, where those countries are not significant food exporters, the war in Ukraine, which is a major grain exporter, is believed to potentially affect food security not only of those who live in the country but people elsewhere in the world. The countries directly involved in the conflict, Ukraine and Russia, contribute to around 15% of global cereal exports, and it is still unknown how the conflict will affect world cereal production and export. In case of reduced export, food security in cereal importing countries might be at risk, if they do not increase domestic production or secure import from other sources in time. Countries such as Sudan, Egypt, Laos, Benin and Somalia import between 75% and 100% of their grain from Ukraine and Russia, and they are particularly vulnerable to this conflict (Buchholz, 2022).

Regardless of whether causes of food shortages are direct or indirect, and range from changes in climate patterns, armed conflicts, growing population, depletion of natural resources to implementation of inappropriate policy instruments, it is highly likely that food shortages in some regions will continue to happen. An FAO report from 2019 suggests that, after years of decline, the number of hungry people has begun to climb again. What's more, the FAO food price index in March 2022 reached its highest level since measurement started in 1961 and is in line with the predictions that we might see more food security episodes in the very near future.

Substantial resources have been dedicated to preventing and remediating food shortages. In 2020, contributions to the WFP alone were around USD 8.5 billion. In March 2021, the Secretary-General of the UN appealed for the emergency mobilization of an additional $5.5 billion to mitigate food security issues fuelled by armed conflicts (UN Security Council, 2021).

Post-famine-period studies are scarce, and interventions are even scarcer. By analysing long-term famine impacts, this dissertation contributes to the literature and illuminates the long-term consequences of food shortages on nutrient intake. If the results of this dissertation reveal long-term effects of famine on nutrient intake, that could be another argument for the international organizations to require more resources.

The example of the Great Famine in China is used as the setting of this research, as it is perceived to be a quasi-natural experiment and exogenous shock to the population affected. An estimated 16.5 million to 33 million people in China died due to hunger during the period of 1959-1961 (Devereux, 2000; Song et al., 2009). *Table A.1* (Annex) presents major famines of the 20th century, and it appears that the famine in China claimed the most human lives. The section below provides a historical perspective of famine episodes in China which preceded the Great Famine and contains background information related to the causes of the Great Famine.

1.1 From the Incredible Famine to The Great Famine

Population in China experienced multiple episodes of famine between 1855 and 1950, and the Table 1 below presents the major disasters and famines in that period.

Table 1: Major disasters and famines 1850-1950

Year	Location	Type	Provinces	Victims (million)	Mortality (est.)
1855	Yellow River	Flood, shift course	1	7	-
1876-79	North China	Drought	5	-	9-13 million
1912	Hai River Basin	Flood	1	1.4	n.e.
1915	Pearl River (Guangdong)	Flood	1-2	6	60,000
1917	Hai River Basin	Flood	1	5.6-5.8	n.e.
1920	North China	Drought	5	30	500,000
1924	Hai River Basin	Flood	1	1.5	n.e.
1928-30	North and Northwest	Drought	8-9	57.3	10 million
1931	Yangzi River	Flood	8-15	61	422,499
1933	Yellow River	Flood	6+	8.2	18,293
1935	Yangzi (mid)	Flood	2	12.7	142,000
1938-47	Yellow River	Flood, shift course	3	6.2	893,303
1939	Hai River Basin	Flood	4	4.5	13,320
1942-43	Henan	Drought, war	1	11-16	2-3 million

Source: Li (2007)

The most lethal drought-famine in imperial China which took place between 1876 and 1879, and which was also known as the Incredible famine, claimed between 9 and 13 million lives. Efforts to mobilize relief around China but also internationally, included “Pictures to Draw Teras from Iron” (Edgerton-Tarpley, 2008). Those illustrations depicted cases of cannibalism, selling women and children, eating tree bark and scenes of suicides caused by famine-induced despair. As we can see, most of those events were nature-initiated disasters such as flood and drought. Having said that, the role of man in those events was also substantial, as large-scale structural changes in agriculture and the associated infrastructural investments, which could have prevented or at least mitigated negative effects of floods and drought were lacking. There are various reasons for the inadequate investments. Almost a permanent state of conflict either with foreign invaders, such as conflicts with Japan in Manchuria and Shanghai, or a civil war between Communist forces and Nationalist forces had devastating impact on property and human casualties, and also prevented long-term structural changes in agriculture. Lack of investments was coupled with numerous taxes that farmers had to pay on both central and local level. Some of the taxes on local level included land tax, poll tax, bandit-suppression duties, military dues and many other (Turvey, 2019). Therefore, the perfect storm which consisted of natural calamities, armed conflicts and tax burden left no room for agriculture to grow. It was only in 1921 when famine prompted China International Famine Relief Commission to introduce reforms in infrastructure, irrigation, wells, cooperation, and credit (Turvey, 2019). Even with those usually short-lived efforts, poverty in certain parts of China has been perceived as a permanent state. R. H. Tawney compared some parts of a rural population with a man who is standing permanently up to the neck in water and even a ripple is sufficient to drown him (Li, 2007).

While both Chinese and international scholars and ideologists recognized and agreed on the dire situation that rural China was experiencing, the solutions to that situation differ greatly. On one hand, western scholars such as John Lossing Buck promoted the idea where family farming would lead the progress and technical improvements in agriculture such as plant and animal breeding would be the pillar of the progress. Additionally, those would be accompanied by institutional infrastructure such as agricultural schools, experimental stations and rural banks (Li, 2007). On the other hand, the Communist movement argued that unequal distribution of wealth and inequality are the main reasons for rural poverty, and that creation of egalitarian society would be a precondition for progress. While many would argue that equality is something that a society should strive for, the mean to achieve the equality that was promoted and implemented by Mao Zedong was debatable. Namely Mao's ideology promoted a radical and revolutionary class conflict, rather than a gradual change (Li, 2007). Overnight land redistribution provided farmers with sufficient amount of land for subsistence agriculture. Additionally, in contrast to corrupt magistrates and other officials as well as merciless militarists who deprived farmers of their grains and animals, communists protected farmers and their crops (Turvey, 2019). In the following period, industrialized centres in and around cities managed to lift some parts of the population out of poverty. On one hand, industry development saw upward trend. On the other, some of the income generated through those industrial centres have been sent back to rural areas and reinvested in improvements in agricultural production. To further promote rapid development of both heavy industry and agriculture, certain structural changes have been pursued. Small and fragmented farms had to be consolidated to increase agricultural output. From 1953 onward, agriculture became collectivized, which enabled some farmers to move to off-farm activities. This shift of labour underpinned the Fist Five-year Plan, which emphasized rapid industrial development. It has been argued that Mao was impatient to see even greater industrial and agricultural gains, which prompted him to launch Great Leap Forward in 1958, and which resulted in creation of large-scale communes in most of rural areas (Li, 2007). The communes opposed any form of individualism. The farmers surrendered their means of production to the commune, they were not allowed to have any plot of land for their own consumption, neither to cook at home. Instead, the "free-food policy" has been adopted, and farmers ate together in canteens "for free". The quantity of foods consumed has been dictated by the canteens, rather than by the workers-consumers. Thanks to very favourable weather conditions, the first year of the Great Leap Forward coincided with a bumper harvest. Hence, farmers could eat as much as they liked, and big amount of food loss and food waste have been recorded (Li, 2007). Encouraged by the results of the first year of the Great Leap Forward, party leaders decided to increase grain procurement, reduce sown area, and reallocate certain share of farmers to non-farm activities (Li, 2007). The following year – 1959 has been marked as the beginning of the Great Famine. It has been widely argued that while grain output in 1959 was lower than in the previous year, it is not food availability but rather food accessibility that caused famine. Table 2 presents some of the factors which have been associated to the Great Famine.

Table 2: The famine and related factors

		1958	1959	1960	1961	1992
1	Famine	Began	Severe	Worst	Milder	Ended
2	Output	High	Low	Low	Low	Rise
3	Weather	Good	Normal	Bad	Bad	Normal
4	Withdrawal rights	No	No	No	No	No
5	Commune size	Extra large	Large	Large	Downsized	Smaller
6	Basic accounting unit	Brigade or even commune	Brigade	Brigade	Was changing to team	Team
7	Work-point reward system	Abolished	No	No	Partially resumed	Yes
8	Private plot	No	No	No	Yes	Yes
9	Resource allocation priority	Heavy industry	Heavy industry	Heavy industry	Agriculture	Agriculture
10	Sown Area	Reasonable	Reduced	Partially recovered	Partially recovered	Partially recovered
11	Net exports	Low	Very high	High	Negative	Negative
12	Procurement	High	Very high	Moderately high	Normal	Normal
13	Communal dining system	Established	Remained	Remained	To close	No

Source: Change and Wen (1997)

Namely, due to political pressure and in order to please political leadership, local cadres overreported grain output, so that the procurement level was much higher than what should have been. Consequently, the amount of food left in canteens to be distributed to farmers was not sufficient. Literature which analysed the relationship between the share of farmers who ate at canteens and famine deaths identified direct relationship between the two (Change and Wen, 1997).

Theory which explores causes of famine through entitlement lenses decouples food availability and famine. In his work on poverty and famines, Sen argues that in private ownership market, entitlement relations include at least four forms: trade-based entitlement, production-based entitlement; own-labour entitlement and inheritance and transfer entitlement (Sen, 1981). Each of the four entitlements can be used for acquiring food. The Great Famine presents the example of what can happen when there is no private ownership market, and when a government controls all four types of entitlements. In other words, when it comes to work environment, ownership and food consumption, the government had an absolute control over the rural dwellers. In urban areas, where communal dining system did not exist, and where residents were allowed to cook at home and plan their consumption themselves the famine largely did not occur (Chang and Wen, 1997).

A separate, yet related work of Amartya Sen, analyses the relationship between a political system and famine. Namely, Sen argues that famine cannot occur where democratic rights and liberties exist (Sen, 2000). The author argues that democratic governments who are subject to public criticism have strong incentives to prevent any catastrophe, including famine, to win elections. In contrast to that, the author positions authoritarian rulers and one-party states, where there is no freedom of speech and free press, and there are no elections to be won or lost. The Great Famine perfectly fits this narrative, as even today, 60 years after the famine Chinese authorities refer to that period as "Three years of natural disasters" and refuse to disclose relevant information which could expose wrong decisions made by the leadership.

1.2 Famine Effects – Short-term and Long-term

Famine very often can induce both short- and long-term consequences. Tan et al. (2014) refer to the culling and scarring effects of famine, where former represents excess mortality during the famine and the latter represents long-term effects not only famine survivors years and decades after the famine, but also those who were in utero during the famine and even children of those who lived through the famine. Fung and Ha (2010) refer to the scarring effect as an echo effect. Some of the scarring effects that have been analysed are BMI, height-for-age, weight-for-age, risks of cardiovascular and metabolic diseases, level of schooling, cognitive ability, labour supply and earnings and mental health. Several main findings are presented in literature review of *Section Two* of this paper. Even food shortages that are less extreme than famine can cause malnutrition manifestations, such as stunting, which has long-term consequences. On one hand, childhood stunting, which represents low height-for-age, results from chronic hunger. As per the World Health Organization, the condition is largely irreversible and has long-term effects on cognitive and physical development, productive capacity and overall health (WHO, 2014). Wasting, that is, low weight-for-height, on the other hand, typically results from a short-term acute hunger episode. Children affected by wasting "have weakened immunity, are susceptible to long term developmental delays, and face an increased risk of death, particularly when wasting is severe" (WHO/UNICEF/WFP, 2014).

Various instruments have been developed to predict, prevent, and address food shortages. While improvements in early warning systems, driven by technology and logistics, can enable international humanitarian aid to be deployed relatively quickly to the areas affected by famine, sometimes that is insufficient or not timely. Either the population of the affected area then suffers or people migrate. The latter is of particular concern for the international community, as famine effects then spill over to neighbouring countries and their populations. Famine by its definition involves hunger, which is recognized by the UN as one of the greatest challenges for achieving global sustainable development. The UN has emphasized the importance of hunger by proclaiming "Ending hunger by 2030" one of its 17 Sustainable Development Goals. Hunger, food shortage, famine and starvation are the terms often used interchangeably. Annex contains definitions of those terms and clarifies differences and similarities between them.

Currently, there are two main groups of instruments that used to prevent or mitigate famine's consequences. In the first group, there are different forms of early warning systems, which help predict famines, warn governments and the international community of a potential disaster, and place food stocks where needed. Another group includes instruments put in place after the event has begun, such as emergency food aid and cash transfers. In emergency situations, the UN's World Food Programme will send in food mainly including wheat flour or rice, lentils and other pulses, vegetable oil, sugar, and salt (World Food Programme, n.d.). WFP's food basket should be sufficient for an individual to survive while the emergency is ongoing. In the aftermath, various national and international supplemental programs will be implemented. Typically, it is poorer quintiles who are mostly affected, as they lack food and financial buffers for extreme events, and they are typically targeted by the food programs. However, it remains to be seen whether individuals who have survived food shortages and escaped the poverty trap also need support.

1.3 Aim, Objectives and Intended Contributions

The main research question which I am aiming to answer in this dissertation is whether people who experienced famine have different nutrient intake comparing to those who have not experienced famine. In other words, I am trying to understand whether famine experience shapes one's life decades after the event, and if it does, whether the effect diminishes over time. Therefore, this dissertation aims to illuminate the long-term consequences of famine on nutrient intake.

To do that, I will examine the impact of exposure to famine on three measures of nutrient intake.

- The first measure is Dietary Diversity Score (DDS), which connects dietary patterns and nutrition outcomes. Examining the relationship between exposure to famine and DDS is the first objective of this dissertation. The current literature established a relationship between DDS and nutrient and micronutrient adequacy, memory, cognitive function, depression, risk of being overweight, metabolic syndrome, cardio metabolic risk factors, risk of fracture and others. Therefore, the results of this dissertation will be relevant to the professionals involved in addressing those conditions.
- The second measure of nutrient intake reflects qualitative and quantitative characteristics of a diet, measured by absolute intake of macronutrients, and relative share of macronutrients in total diet. Those macronutrients are fat, protein and carbohydrates. Examining the relationship between exposure to famine and qualitative and quantitative characteristics of a diet is the second objective of this dissertation.
- The third measure of nutrient intake is measured by household food expenditures. Examining the relationship between exposure to famine and household food expenditures is the third objective of this dissertation. As current literature posits, food expenditures are directly proportional to quality of diet. Therefore, the results of this dissertation could bring more evidence of diet quality determinants

<u>Contribution</u>

Nutrient intake of an individual is influenced by many different factors. Some of these factors are well explored and understood: for example, personal and household income, educational attainment, urbanization, cultural and social norms, immediate as well as wider food environment, policy measures and others.

Other factors, such as the long-term impacts of food shortage on nutrient intake have received less, if any attention in the existing literature. By exploring this relationship, my study will contribute to the body of literature that is investigating underlying determinants of nutrient intake, and bring new perspective on the long-term effects of famine episodes.

Furthermore, if my assumption that famine has long-term negative impact on nutrient intake proves to be correct, my results would support work of health care workers, nutritionists, and policy makers. Nutritionists would better understand why some sub-populations, such as famine survivors, have poorer nutrient intake and could tailor their approach accordingly. Health care workers would be able to detect certain medical conditions earlier and prevent larger healthcare expenditures caused by the advanced phase of those conditions. Finally, policymakers in consultations with nutritionists and health care workers would be able to create targeted policy instrument to provide long-term support to famine survivors. This could translate to improved public health and increased overall labour productivity.

1.4 Organization of the Dissertation

This dissertation is organized as follows:

In ***Section 2***, I explore impact of exposure to famine on Dietary Diversity Score (DDS), which will be proxy measure for nutrient intake. I analyse whether those who experienced famine or live in famine affected areas, eat more or less diverse foods from those who were not affected.

In ***Section 3*** of the dissertation, I explore the relationship between famine and diet quality, measured by share of macronutrients in total diet. More specifically, I examine whether relative share of fat, protein and carbohydrates in total diet is different from those who were directly or indirectly exposed to famine comparing to non-exposed parts of Chinese population.

In ***Section 4***, I explore the relationship between famine and nutrient intake, measured by household expenditures and family planning decisions. More particularly, I test if food shortage episode and exposure to the Great Famine in childhood affect food expenditure later in life. I also test if exposure to famine affects other lifestyle expenditures, such as non-work travel expenditures, entertainment, and charitable donations expenditures. Finally, in this section I also analyse impact of exposure and degree of famine on those family planning decisions, which have nutritional consequences.

Section 5 connects the results from sections two, three, and four. This section aims to capture and synthesize all main findings from the dissertation in a coherent way.

In ***Section 6*** I further elaborate the effects of famine on nutrient intake and introduce a policy dimension to the dissertation. Namely I propose policy instruments which could reduce the negative effects of famine on nutrient intake.

Section 7 contains summary of my findings and presents limitations of my dissertation. Additionally, in this section I propose ideas and directions for further research. Some of the ideas are directly related to my dissertation and could present continuation of my work. Others reveal new aspects of analysis of famine long-term effects that are not covered by the existing literature.

2 Long-term Effects of Famine on Dietary Diversity Score

Here, I explore impact of exposure to famine on nutrient intake, where dietary diversity score (DDS) is proxy measure for nutrient intake. DDS is a measure which connects dietary patterns and nutritional outcomes. I analyse whether those who experienced famine or live in famine affected areas, eat more or less diverse foods from those who were not affected. In addition to this, I look at factors which could potentially have mitigating effects of famine, and those factors are income, education, age and gender.

2.1 Evidence from the Existing Literature

The literature extensively covers the long-term effects of extreme events, such as famine, armed conflict, natural disasters, etc. The public health literature examines the long-term effects of extreme events on body mass index (BMI), life expectancy, post-traumatic stress reactivity and susceptibility to cardiovascular diseases and other medical conditions. The economics literature covers long-term effects of extreme events on employment, schooling, income later in life, risk-taking behaviour, and savings decisions. This dissertation aims to connect the public health and economics dimensions by studying effects of an extreme event on nutrient intake, which drive medical and economic outcomes. *Section Two* and *Section Three* explore DDS, while *Section Four* examines lifestyle choices. All three sections reveal a dimension of the long-term effects of extreme events not covered by the current literature and provide guidance for policy recommendations.

"Dietary patterns" and "food choices" are terms used interchangeably in the literature. While the two are related, for the purpose of this dissertation, food choice will denote short-term action involving single food items, while dietary patterns will represent aggregated food choices over a longer period or a combination of food items. Both terms are elaborated on in the paragraphs below.

The literature which examines factors influencing food choices has mainly focused on the immediate and underlying determinants which are concurrent with food consumption. There are various classifications of these determinants of food choice. Turner et al. (2018) present the food environment and its parts as determinants of consumption. The authors divide the food environment into two domains: external and personal. In the external domain, all factors are divided into four groups: food availability, prices, vendor and product properties and marketing and regulation. In the personal domain, food consumption determinants are also divided into four groups: accessibility, affordability, convenience and desirability. Furthermore, the EUFIC (2006) divides the major food choice determinants in six groups. Some groups are similar to Turner's factors, and these are the economic, physical and social determinants. Other EUFIC groups relate to an individual's characteristics and needs. These are biological determinants (hunger, appetite, and taste), psychological determinants (mood, stress and guilt) and, in the last group, attitudes, beliefs and knowledge about food. The literature over time has slightly shifted from examining the direct causes of food choice to examining underlying causes. The reason for this might be that despite of immense resources and efforts by international organizations, governments, and civil sector organizations to address malnutrition and the associated conditions, very modest progress has been made. The Nutrition Equity Framework is one of the tools that helps explain the role of socio-political context in the social stratification that leads to inequitable food, care, and health environments and finally to malnutrition (Nisbett et al., 2022).

Food choice analysed from a broader and more aggregated view leads to an analysis of dietary patterns. There is no single approach to identify these patterns. The process depends on the research question or data availability.

Share of carbohydrates in total diet is one critical measure. As per nutrition transition trajectory, a society typically moves from cereal-based diets during a famine era, to a more diverse diet, which includes more fruits, vegetables, and animal-sourced products, and then to a diet that contains more fat, sugar and processed foods. It finishes its "evolution" with an increased share of carbohydrates, fruits and vegetables and a decreased share of fat and processed foods (Popkin, 1993). Hence, observing macronutrients only, including their share in the total diet can also indicate the main characteristics of one's diet. The macronutrients are carbohydrates, proteins, and fat.

Another approach to dietary pattern analysis is based on principal component analysis. By using this approach, study sample is grouped according to the similarity in food products and food groups consumed by the sample population. Some of the diets in China identified using this approach are "Traditional Southern" (high intake of rice, pork and vegetables), "Snack" (fruits, eggs and cakes), "Western" (high intake of poultry, fast foods and milk) (Zhang et al., 2016), "Meat" (animal products and alcohol), "Healthy" (fruits, milk and vegetables), "High-energy" (high-energy food), "Traditional" (cereals, potatoes, beans and vegetables) (Lyu et al., 2014), "Traditional Chinese" (rice, vegetables, poultry, pork and fish) and "Modern" (wheat, processed meats and fast food) (Zhen et al., 2018). Finally, a family of multiple indicators measures both dietary patterns and the predisposition for nutritional and health outcomes. Those outcomes are stunting, wasting, obesity, CVD, etc.

Dietary Diversity Score (DDS) is an indicator that connects dietary patterns and nutritional outcomes. It is established that DDS is positively associated with nutrient adequacy, and child growth (Ruel, 2003). In technical terms, DDS presents the number of food groups that an individual consumed within a defined period and is based on self-reported information. The number of food groups varies across questionnaires, from five to 24, and this number depends on the level of food products aggregated. In the case of a high level of aggregation, the groups can be grains; fruits and vegetables; meat, poultry, and seafood; dairy; and beans, eggs and nuts. A higher level of disaggregation would split fruits and vegetables into vitamin A rich fruits, other fruits, vitamin A rich vegetables and tubers, dark leafy vegetables, and other vegetables.

The most widely used timeframe in the questionnaire is a three-day or 24h recall period. Without disclosing the quantities consumed, as long as they are above a certain threshold, respondents are asked to report what they consumed in a given timeframe. The recently developed country-adapted diet quality questionnaire (DQQ) contains a set of predefined questions about consumption, where respondents answer with yes/no (Global Diet Quality Project, 2021). A DDS can be based on a different set of questions and different number of food groups, depending on the main purpose of the questionnaire. Food groups and questions in the research analysing micronutrient deficiency can differ from those in research investigating obesity and noncommunicable diseases (NCDs). Furthermore, depending on the unit of analysis, individual or household, different approaches are used. The literature explores factors that influence DDS and the relationship between DDS and various nutrition- and health-related outcomes. Some of those outcomes are nutrient and micronutrient adequacy, memory, cognitive function, depression, risk of being overweight, metabolic syndrome, cardio metabolic risk factors, risk of fracture and others.

The findings suggest that DDS is positively associated with nutrient and micronutrient adequacy (Wang et al., 2021; Zhang et al., 2017; W. Zhao et al., 2017) and memory and cognitive abilities in various age groups (Li et al., 2021; Yin et al., 2017; Zhang and Zhao 2020; Zhang and Zhao, 2021), while the association with BMI, obesity and metabolic syndrome is inconclusive or does not show significance (Qorbani et al., 2021; Salehi-Abargouei et al., 2016). Furthermore, in some cases, the relationship is not linear. The association with metabolic syndrome is positive for a younger population and negative for an older one (Tian et al., 2017), or in some cases the effect diminishes over time (Zhang et al., 2020). Finally, research with regression results split by DDS levels finds a negative correlation between DDS and incidence of being overweight and obesity in subsamples consuming more than six out of nine food groups, emphasizing the importance of understanding each additional food group consumed and suggesting that consumption of an additional food group might lower the intake of high-fat and high-calorie foods (Tao et al., 2020). Several studies explore both DDS and food variety within a single food group and find that they also affect health outcomes

(Conklin et al., 2016; Tian et al., 2017). The literature that explores the association between DDS and different health conditions mainly relies on cross-sectional data, and that is one of its main shortcomings, as health outcomes are often affected not only by observable but also by unobservable characteristics. The use of panel data is important in such studies.

Factors that affect DDS

DDS both influences and is influenced by various factors. The literature mainly explores how factors such as age, location of residence, education, income, alcohol consumption and smoking habits, ethnicity, marital status, proximity to different types of markets, costs associated with access to diversified food and nutritional and dietary knowledge affect DDS. Having said that, those covariates are mainly used as controls, and details on how or why they affect DDS is scarce.

Household income and food consumption has been widely studied. In 1857, Ernst Engel identified an inverse relationship between household income and the relative share of food expenditures. Furthermore, the nutrition transition, as first described by Barry Popkin, involves the evolution of societies living in relative poverty and relying on cereal-based diets to more affluent societies, consuming more diverse diets (Popkin, 1993). The relationship between income and DDS has been shown to be positive. Yet the strength of the relationship depends on income terciles, where the relationship is stronger in the second tercile than in the third tercile (Hou et al., 2021). The difference in DDS based on the different levels and the diminishing impact of income along the income distribution has been confirmed by several authors (Doan, 2014; Wang et al., 2021; Zhang et al., 2017; Zhao et al., 2020).

Another household characteristic that has been found to positively affect DDS is education (Bi et al., 2019; Hou et al., 2021; Zhong et al., 2018), and this relationship holds irrespective of whether the education level is of a respondent or respondent's parents in case of school age children. Furthermore, a positive association between education and a dietary knowledge score has been established (Xu et al., 2020). Finally, a positive relationship between nutritional knowledge and DDS has been found (Wang et al., 2021), suggesting that dietary knowledge might be a channel through which education affects DDS.

Comparing to minority nationalities in China, members of the *Han nationality* have more diverse diet, their share of cereals is smaller, and their protein intake is higher (Ge et al., 1997). The authors argue that the reasons that national minorities in China have a poorer diet, could be related to their residence and income status and educational attainment. They argue that minority populations typically live in rural areas and have less income and education, relative to the Han majority. Zhang et al. (2017) also finds that the Han have a higher DDS, but the authors argue that a higher DDS is not automatically better than a lower one, as high DDS can lead to overconsumption and obesity.

In addition to education, income, and Han ethnicity, other factors such as urban residence, marital status and nutritional and dietary knowledge have been positively associated with DDS, while smoking and alcohol consumption have been negatively associated with it (Hou et al., 2021; Wang et al., 2021; Zhang, et al., 2020; Zhang and Zhao 2021; Zhang et al., 2017; Zhao et al., 2020; Zhao et al., 2017; Zhong et al., 2018). Findings related to impact of age and gender on DDS are inconclusive.

What is affected by extreme events?

Research that addresses the long-term effects of extreme events does not explore the relationship with dietary patterns. Instead, the following personal characteristics have been found to be affected by extreme events: risk taking behaviour (Ben-Zur and Zeidner, 2009; Bernile et al., 2017; Bucciol and Zarri, 2015; Kim and Lee, 2014), altruism (Y. Li et al., 2013), accounting conservatism and financial disclosure quality (Hu et al., 2017; Yao et al., 2020), overconfidence (Malmendier et al., 2011) and later in life vulnerability and growth (Updegraff and Taylor, 2000). The literature shows that that an individual who experiences an extreme event tends to have a higher savings rate (Chen et al., 2018), perceive him- or herself as facing a greater risk of future disasters (Cameron and Shah, 2015) and show increased preparedness for natural disasters and hazards, which increases with the severity of past damage (Weinstein, 1989). Additionally, the research examining the effects of the Great Famine finds that CEOs who experienced it use more conservative accounting (Hu et al., 2017), and that they use less debt, hold more cash and pursue fewer yet better-performing takeovers (Zhang, 2017), while managers who lived through the famine are less like to behave unethically and their firms show better performance during an economic downturn (Feng and Johansson, 2018; Yao et al., 2020). The only paper that explores the relationship between exposure to famine and "food related behaviour" found that there is negative correlation between famine experience and food wasting behaviour (Ding et al., 2022).

Identifying cohorts

The literature identifies age as one of the main determinants of the long-term effect of famine, and in estimations of the effects of famine on dietary, nutrition and health outcomes, research controls for respondents' age. However, very often the relationship is not linear. Therefore, authors have examined different age and birth-year cohorts, where the definition of the cohort is often arbitrary.

In case of Great Famine survivors, some authors have restricted the analysis to those born between 1956 and 1964, a relatively narrow interval, to increase the similarity of the unobserved later-life factors (Almond et al., 2010). Others follow the same route and divide the sample into three **birth year cohorts**: pre-famine (1956–1958), famine (1959–1961) and post-famine (1962–1964)(Xu et al., 2016). Xu et al. (2016) use their three cohorts to obtain a famine severity index. Yet for the purpose of their statistical model, they keep only the famine and post-famine cohorts. Li et al. (2010) divide their sample into five cohorts: non-exposed (1 Oct. 1962 to 30 to Sept. 1964), foetal-exposed (1 Oct. 1959 to 30 Sept. 1961), early-childhood exposed (1 Oct. 1956 to 30 Sept. 1958), mid-childhood exposed (1 Oct. 1954 to 30 Sept. 1956) and late childhood exposed (1 Oct. 1952 to 30 Sept. 1954). Rong et al. (2019) apply a very similar approach, dividing their sample into five cohorts and relying only on years of birth. Ning et al. (2019) similarly have foetal exposure (1 Jan. 1959 to 31 Dec. 1961), childhood exposure (1 Jan. 1949 to 31 Dec. 1958) and adolescent/adult exposure (1 Jan. 1931 to 31 Dec. 1948), while considering those born between 1 Jan. 1962 and 1 July 1975 as an unexposed cohort. Hu et al. (2017) use the Urban and Rural Health Survey in their analysis of long-term effects of the Great Famine and divide the sample into following cohorts: 1939–1942, 1943–1958, 1959–1961 and 1963–1965. Tan et al. (2014) explore the effects of famine on different health characteristics of children born to parents who were born from 1963–1965 (control group) and children born to parents who were born from 1959–1961.

In a studying of the Leningrad siege starvation and its impacts to health, the **age cohorts** were less than eight years old, nine to 15 years and 16 to 26 years (Sparen et al., 2004). The Ethiopian Great Famine has also been the subject of empirical analysis. In a study examining its effect on cognitive function of different age cohorts, the authors divided the sample into postnatal exposure, prenatal exposure and no exposure (Arage et al., 2020). In research that analyses the effects of famine on risk taking later in life, Cheng and Smyth (2021) explore the effect on five birth cohorts, which were based on an age-related child development stage, relative to famine years. Those are the unborn (1965–1968), unborn (1962–1964), infancy (1959–1961), early childhood (1956–1958) and middle to late childhood (1944–1955). The medical and psychiatric literatures focus on critical **developmental phases** and the influence of the environment during those phases on long-term effects. Medical researchers concentrate on the phases relevant for their research question. Sometimes those are prenatal, infant, early childhood and adolescent periods when examining life-course perspective (Herman et al., 2014), or childhood, adolescence, young adulthood, midlife and older adulthood in examining post-traumatic stress disorder (Ogle et al., 2013). Sometimes the periods are narrower, and this is commonly used in the literature focusing on critical phases of neurocognitive development. In this case, the cohorts are first trimester, early-mid gestation and mid-late gestation (Irwin et al., 2016) or gestation phase and early postnatal stage (Benton, 2010).

Heterogenous effects by cohort

When it comes to effects of extreme events early in life on later consequences, Almond et al. (2010) find that exposure to famine in utero has the most profound effects, relative to those born before or afterward. Contrary to that, Xu et al. (2016) find little evidence supporting the foetal origin hypothesis, and they find that, because of mortality selection, prenatal exposure reduces disease risks later in life. Li et al. (2010) find that, relative to those born after the famine, both foetal- and childhood-exposed cohorts were shorter. Yet those exposed during their childhood were most severely affected. Sparen et al. (2004) find that systolic blood pressure was particularly high among those who were around puberty (nine to 15 years) during the siege of Leningrad and experienced severe food shortage. Qin et al. (2020) explore the relationship between exposure to famine of different age groups and metabolic syndrome (MetS) and find that individuals exposed to famine during the foetal development and childhood had a higher risk of metabolic syndrome. Ning et al. (2019) have also studied the relationship between exposure to famine and MetS, and their findings are opposite of those of Qin et al. (2020). They suggest that, when comparing foetal, childhood and adolescence/adult age groups, the highest prevalence rate of MetS was among the oldest cohort, and the smallest prevalence among youngest—the foetal cohort. Similarly, Hu et al. (2017) find that individuals exposed to famine in the prenatal/infant stage have reduced risks of chronic disease in later life, while those exposed to famine in childhood/puberty have an increased risk of chronic disease. This might be explained by the positive selection effect, which would lead to only the strongest individuals surviving the famine.

In literature that analyses impacts of early life adverse events on behaviour, Cheng and Smyth (2021) find that, relative to those born after a famine, infancy and early childhood cohorts show less risky behaviour, while, among those who hold risky assets, the ones exposed in middle to late childhood show riskier behaviour. The life-course literature posits that every stage from prenatal to adolescence is critical for development of cognitive, emotional, social and physical abilities, and hence a suboptimal environment in different phases can have negative long-term effects on those abilities (Herman et al., 2014). In addition, the complexity of analysing early life influences on later-life effects comes from the fact that different dimensions of the influences must be considered. In examining sensitive periods for exposure to adversity, a comparison was made between recency model, an accumulation model and a sensitive period model (Dunn et al. (2019) in Gabard-Durnam and McLaughlin, 2019). The recency model analyses whether recent experiences are most influential, the accumulation model looks at the relationship between number of occurrences and the effect, and the sensitive period model investigates which developmental period is the most influential. Dunn et al. (2019) find that timing the of exposure best explains the relationship between an adverse impact before three years of age and effects later in life, at least when it comes to influence on brain development and associated long-term effects on emotions, cognition and behaviour. In the famine-related literature, it has been argued that memories of famine peak between the ages of five and seven and then plateau (Cui et al., 2020; van den Berg et al., 2016).

Defining famine severity

To directly determine famine severity, the literature applies different approaches, which fall under two main categories: excess death rate in time of famine and the difference in cohort size in famine and non-famine years derived from 1990 Population Census. Xu et al. (2016) employ two measures of famine severity. The first one, excess death rate (EDR), presents the difference between mortality rate in famine years (1959–1961) and the average death rate three years before the famine (1956–1958). As the authors point out, this approach has two shortcomings: the death rates were published by the State Statistical Bureau and were thus subject to data distortion, and the second limitation is that these were provincial mortality rates which could not capture sub-provincial differences. The second measure that Xu et al. (2016) apply is called cohort size shrinkage indices (CSSI), and it uses 1990 China Population Census. The authors use the average cohort size of those born three years after and three years before the famine (non-famine years) and compare to the cohort size of those born in famine years (1959–1961). To do that, they subtract the famine cohort from the non-famine cohort and divide by the non-famine cohort, so that a larger value of the index indicates a greater reduction in cohort size due to reduced fertility and increased infant mortality, presumably caused by the famine. The authors further argue that both EDR and CSSI could capture famine severity, should several assumptions hold. These are accuracy of the census data on mortality and fertility, stability of the secular trend in mortality and fertility absent the famine and restriction of population migration. Similarly, Meng et al. (2015) use 1990 Census to count the number of missing people from the 1959–1961 cohort. Li et al. (2010) use EDR in their estimations. They calculate the percentage change in mortality rate between the mean level in 1956–1958 to the highest value during the period of 1959–1961. The regions with EDR of 50% and higher are categorized as severely affected famine areas and those below 50% as less severely affected. To perform sensitivity analyses, the authors apply more conservative EDR cut-offs. That is, they use EDR 100% and higher to define severely affected famine areas. Li et al. (2011) also use an EDR of 100% as a cut-off to define severely affected areas.

Xu et al. (2016) adopt an instrumental variables (IV) approach to assess famine severity. They rely on evidence that local cadres exaggerated grain yields to be rewarded with career advancement, salary prestige and other privileges, which contributed to the severity of the famine. Therefore, the authors argue that the frequency of exaggerating grain yields (IV) should relate positively to famine severity in a given prefecture. The IV was constructed by using newspaper reports of exaggerations of grain yields by county officials in each prefecture in the *People's Daily* archives.

2.2 Hypothesis Development

While the literature has found evidence of economic, medical, cognitive, and behavioural consequences of famine, as described above, the enduring effects of famine on dietary diversity and food consumption have received less attention. This paper aims to fill this gap.

Existing evidence suggests that, in cases of extreme events other than famine, education, income, Han ethnicity, urban residence, marital status and nutritional and dietary knowledge have been positively associated with the dietary diversity score, as a proxy for the food consumption patterns, while smoking and alcohol consumption have been negatively associated with it (Hou et al., 2021; Wang et al., 2021; Zhang et al., 2020; Zhang and Zhao, 2021; Zhang et al., 2017; Zhao et al., 2020; Zhao et al., 2017; Zhong et al., 2018).

The aim of this dissertation is to study the effects of famine on nutrient intake later in life. In this *Section*, I use Dietary Diversity Score (DDS) as a proxy for food consumption patterns. In subsequent *Sections*, I augment my analysis to study macro-nutrient intake as well as lifestyle choices, such as eating out expenditures, alcohol and tobacco intake, travel and leisure spending, and family planning decisions.

Based on the literature (Chen et al., 2018; Ding et al., 2022; Zhang, 2017), which explores behaviour of individuals who survived extreme adversity early in their lives and argues that they are more likely to show more caution and frugality years after the event, my first hypothesis is:

H1: People who experienced famine have lower dietary diversity score relative to those who have not.

The literature also suggests that there are potentially heterogenous effects of early life adverse events on later decision-making. Yet there is little conclusive evidence of this effects on DDS. The research (Cui et al., 2020; Dunn et al., 2019; van den Berg et al., 2016) that is closest to what will be analysed in this dissertation argues that it is very likely that survivors who were between ages of five and seven during the famine have the most intense memories. This motivates my second hypothesis:

H2: Age at which an individual was exposed to an adverse life event (such as famine) plays a significant role in determining the strength of its effect on dietary diversity.

Additionally, the literature shows that household income has a positive impact on DDS, yet the impact shows a diminishing effect along the income distribution. It has a more important role in lower income quintiles. The impact of income in famine-affected areas has not been explored, and so, based on the literature, my next hypothesis is as follows.

H3: *Among famine survivors, higher income mitigates negative effects of famine on dietary diversity score and within subpopulation in the same income brackets the effect of famine on DDS is driven by famine severity.*

A positive association between education and DDS has been widely recognized in the literature (Drescher et al., 2009; Moon et al., 2002; Variyam et al., 1998), and the established channel for this association is dietary knowledge. Here I argue that this relationship is equally important when an adverse event, such as famine, happens and that, among affected individuals, there is a positive relationship between education and DDS. Thus, my hypothesis is:

H4: Among famine survivors, more education mitigates negative effects of famine on dietary diversity score, and within subpopulations of the same education level, the effect of famine on DDS is driven by famine severity.

As research does not provide conclusive evidence on impact of gender on DDS, my hypothesis is that the famine has heterogenous effect by gender. Thus, my hypothesis is:

H5: Among famine survivors, men and women are not equally affected.

2.3 Data

I use several different data sources: the *China Health and Nutrition Survey* (CHNS) as well as death rates as used by Meng et al. (2015), which are based on *China National Bureau of Statistics* (NBS).

CHNS's dynamic cohort data is a joint work of social and biomedical scientists from the Carolina Population Center at the University of North Carolina at Chapel Hill (UNC-CH) and the National Institute for Nutrition and Health of the Chinese Center for Disease Control and Prevention (CCDCP). The survey was primarily designed to assess the impacts of economic and social transformations on the health and nutrition of the Chinese population in the post trade-liberalization period. To that end, a set of household and individual socioeconomic and demographic factors have been collected.

The pilot survey was conducted in 1989 and included 3,795 households from 180 communities and eight provinces. Those provinces are Liaoning, Jiangsu, Shandong, Henan, Hubei, Hunan, Guangxi and Guizhou. The last wave was surveyed in 2015, and it included 7,319 households from 388 communities, from the following 15 provinces and municipalities: Beijing, Chongqing, Guangxi, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning, Shaanxi, Shandong, Shanghai, Yunnan and Zhejiang. More details about the number of households and provinces per wave can be found in Annex 1, Table A.1.1.

China has three-tier administrative system. Tier 1 are provinces, autonomous regions and municipalities under the central government. Tier 2 includes provinces and autonomous regions divided into autonomous prefectures, counties, autonomous counties and cities. Tier 3 encompasses counties or autonomous counties subdivided into townships, ethnic townships, and towns. In CHNS, **Tier 1** units are provinces and municipal cities, **Tier 2** are urban/rural areas, **Tier 3** are cities and counties, and **Tier 4** are urban/suburban neighbourhoods (in the case of cities) and villages (in the case of counties). The primary sampling unit is a community, and, in the last wave, there were: 60 urban neighbourhoods, 60 suburban neighbourhoods, 30 towns, and 180 villages.

The sample was chosen through a multistage random cluster process, which covered counties and cities stratified by income (low, middle and high). Furthermore, four counties and two cities were randomly selected in each province. Urban and suburban neighbourhoods within cities and townships and villages within the counties were selected randomly. Those four categories—urban and suburban neighbourhoods, townships and villages—are communities. Finally, 20 households were randomly chosen in each community.

The number of communities and households differs across the waves. The baseline survey (1989) included 3,795 households, while the last available survey (2015) included 7,319 households. Some of the 15 provinces were part of CHNS from the beginning of the project until the last wave, some were added later, and some took part at the beginning of the project, withdrew and later returned to the project. The survey years were 1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009, 2011 and 2015. In 1989, health and nutrition data were collected only from pre-schoolers and those 20 to 45 years old. Since 1991, all household members have been interviewed.

The dataset includes the following elements.

The *Household Survey* contains information about the main socio-demographic household characteristics, such as income and wealth, housing characteristics, labour force participation and medical insurance. Health information also includes details on time and money associated with curative and preventive care. There is also information about age, gender, education, marital status and ethnicity of the household head. Particular attention has been paid to income. In addition to income from market and nonmarket activities, the survey included nonmonetary government subsidies.

The Health and Nutrition Survey contains individual data from detailed physical examinations, such as blood pressure, body composition, health-related habits and history (smoking, alcohol consumption, medication use and chronic disease) and individual dietary data. The survey has been implemented by trained nutritionists, who had experience with national surveys. Household-level food intake data were recorded during three randomly selected consecutive days, by comparing food quantities before and after meals and considering the number of people who consumed food. This is important, as a household's guests would also be counted. Food and preparation waste was also recorded, including spoiled rice and food used as pet feed. Individual data were collected by applying 24-h recall data for three consecutive days, and the questionnaire included foods consumed at home and away from home. Individual and household data would then be compared, and any large discrepancies would be addressed. This section of the dataset contains not only information about dietary intake but also about food preparation methods, health risks behaviour such as smoking and alcohol consumption, key chronic diseases and anthropometric information. There are three different Food Composition Tables (FCT) used throughout the survey. Survey years 1989, 1991 and 1993 used 1981 FCT; survey years 1997 and 2000 used 1991 FCT; survey years 2004, 2006, 2009 and 2011 used 2002/2004 FCT. The main purpose of the FCT is to facilitate conversion of food items into nutrients. To increase the sample size, FCT 1991 and 2002/2004 have been merged, and more details are provided in methodology section.

Physical Activity information has been collected both for children and adults. Data collected from children can be grouped into questionnaire-based data and Caltrac Actometer data. The first group provided insight into habits related to physical and sedentary activities, both in and outside of school Some of sedentary-related questions were used to identify habits related to watching TV, computer use, reading, playing board games, etc. The second group involved use of a device, which measures total energy expenditure, and the measurement data were complemented by a set of questions related to sleeping and nonsleeping activities.

Adults responded to questions about their daily physical activity, mainly relating to intensity of their work.

The *Elderly Component* reflects not only activities of daily living (ADL) but also cognition. Additionally, it includes questions about inter-generational transfers, which reflect exchanges of food, clothing, childcare and elder care. When there was care for elders, the questionnaire included information about both caregiver and caretaker, their relationship, age, gender, occupation, and political status.

The *Body Image and Mass Media Behaviours and Practices* section includes questions about desired and actual body type as well as mass media and television related questions. Mass media questions captured not only how frequently children watched TV and whether they ate or snacked while watching, but they also revealed intrahousehold dynamics by asking questions about watching TV with one or both parents, communication with parents while watching, including asking parents to buy products advertised on TV, etc.

The *Ever-Married Women Survey* contains data on family planning, infant feeding practices and pregnancies, marriage and fertility history.

The *Community Survey* questionnaire has been shared with a knowledgeable respondent about community infrastructure, services, prevailing wages, and population. One questionnaire was completed for each community. While the CHNS individual and household dataset is freely available on https://www.cpc.unc.edu/projects/china, the community-level dataset is the only part of CHNS that is subject to the project director approval.

The *Food Market Survey* has changed over time. The first two waves included both state and free-market prices, while the subsequent waves contain only information from free-market stores. The food market survey included dominant price information about commonly consumed food products, such as rice, bleached and unbleached flour, noodles, millet, refined oil, various vegetables and milk.

The Health and Family Planning Facility Surveys reflect the state of the healthcare facilities, personnel, sources of funds, prices of available services, and distance to the primary sampling unit. This section also contains information about local family planning policies

In addition to the survey elements above, only one wave - 2009, contains three additional surveys: blood collection, toenail collection and boy maturation survey. The blood and toenail samples were used to obtain biomarker measurements. Some of the biomarkers collected were albumin, creatinine, glucose, high-density lipoprotein cholesterol, insulin, triglyceride, urea, and haemoglobin.

Summary statistics of the covariates used in this chapter are presented in the *Methodology and empirical strategy* part of this *Section*.

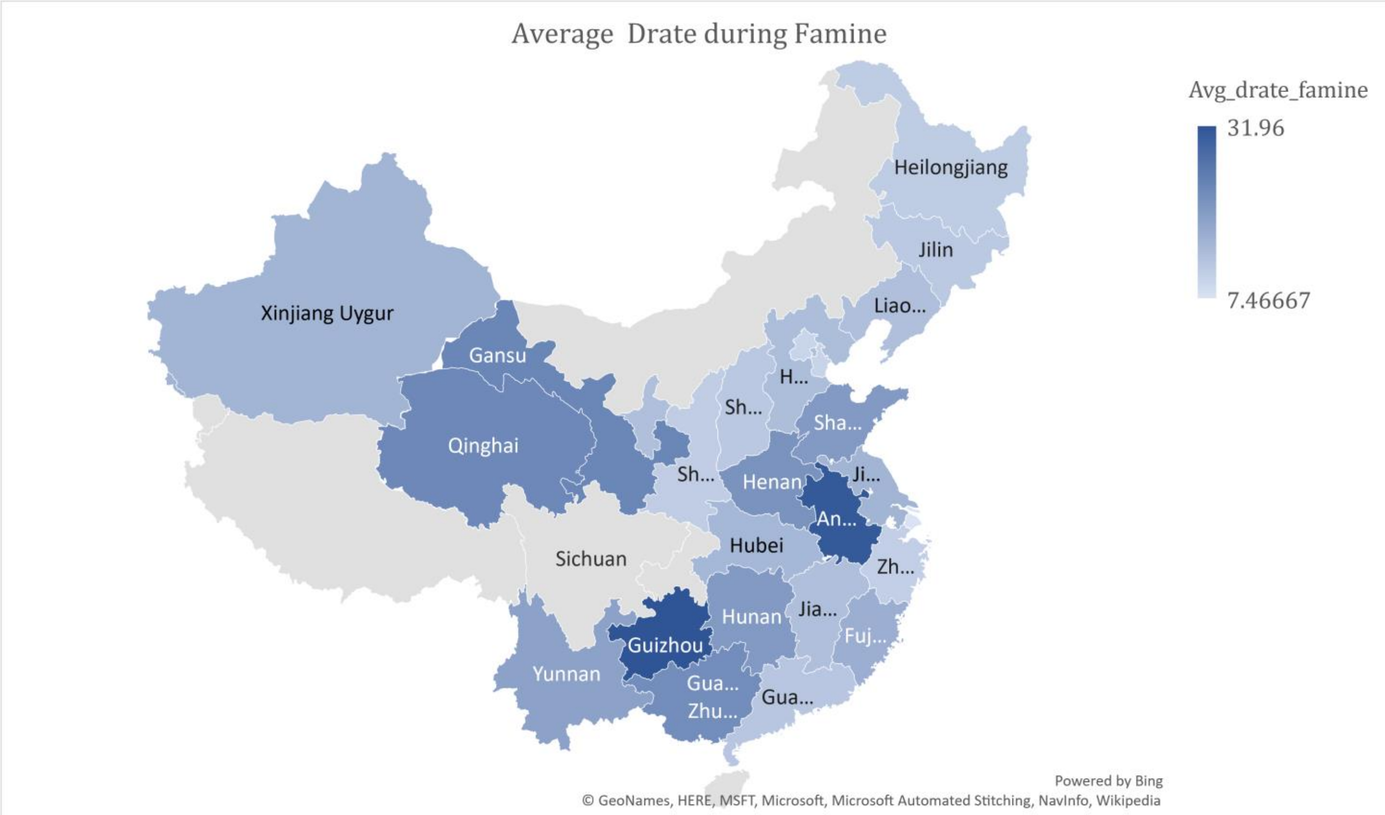
The map of regions that participated in CHNS is presented below (Map 1.1).

In addition to CHNS, I also use the supplementary materials of Meng et al. (2015), who provide information on death rates in all provinces in China. This dataset is used for calculation of the excess death rate. Death rates in all provinces are presented below, while graphs with separate provinces are provided in Annex 1 (Graph A.1.1 – A.1.11)

Map 1. 1: CHNS Survey regions

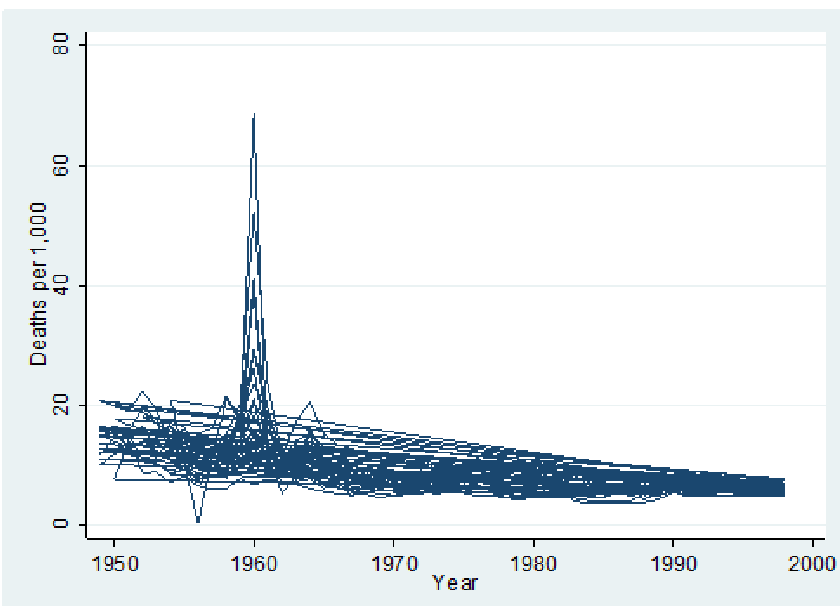
Source: CHNS

Map 1. 2: Average Death Rate during famine years (per 1000)



Source: Meng et al. (2015)

Graph 1. 1: Death rate (Deaths per 1000) across provinces in China 1949-1998



Source: Meng et al. (2015)

2.4 Methodology and Empirical Strategy

Methodology

In the "*Literature review*" section, I presented what has been covered by the studies which examine either extreme events, or dietary patterns or both, and I also pointed at the research gaps, which this dissertation aims to fill. In this section, I provide a detailed description of variable construction and describe my empirical strategy.

Dietary patterns

The main variable of interest is nutrient, which is being proxied by **Dietary Diversity Score (DDS)**. In the "*Literature review*" part (sub-section 2.1), I elaborate what DDS is, its main characteristics, and how current literature uses it in the analysis. As detailed, the number of food groups in DDS varies across literature, from 5 to 24, and this number depends on the level of food products aggregation. Additionally, DDS can be based on a different set of questions and different number of food groups, depending on the main purpose of the questionnaire.

DDS is obtained from CHNS dataset, which contains information about food consumption on individual and household level. All food consumed is being coded using China food composition tables (CFCT) in the format XX-X-XXX. Stata code "11101" corresponds to CFCT code "01-1-101". The first two digits of CFCT denote a food group and based on that all foods are being divided into 20 food groups.

Those 20 groups are:

- Cereals and Cereal Products
- Tubers, Starches and Products
- Dried Legumes and Legume Products
- Vegetable and Vegetable Products
- Fungi and Algae
- Fruit and Fruit Products
- Nuts and Seeds
- Meat and Meat Products
- Poultry and Poultry Products
- Milk and Milk Products

- Eggs and Egg Products
- Fish, Shellfish, Molluscs
- Infant Foods
- Ethnic Food and Cakes
- Fast Foods
- Beverages
- Liquor and Alcoholic Beverages
- Sugars and Preserves
- Fats and Oils
- Condiments

Some of the existing studies which uses DDS to examine health outcomes sets a minimum quantity threshold, so that only food consumed in excess of 2 or 10 or 15g are taken into account. In CHNS, minimum and maximum threshold for household consumption were 20 grams and 15 kilograms. The DDS in my analysis presents ratio between consumed food groups and total number of food groups, where minimum is 1/20 (0.05) and maximum is 20/20 (1.00). Consumption of 5 food groups would correspond to score 5/20 (0.25).

Current literature uses various number of food groups in their analysis. This number varies, and there is no single approach to this. Instead, the choice of types and number of food groups depend on the research question. In this paper I start by using 20 food groups described above and based on the China Food Composition Tables 2002/2004. To verify if the results are sensitive to different number of food groups, I first merged the original 20 groups to 12 groups and then to 6 food groups. The 12 groups followed FAO Household DDS methodology, which is based on Food and Technical Assistance Project (FANTA).

Those 12 groups are:

- Cereals
- White roots and tubers
- Vegetables
- Fruits
- Meat
- Eggs
- Fish and sea food

- Legumes, nuts, and seeds
- Milk and milk products
- Oils and Fats
- Sweets
- Spices, condiments, beverages

Finally, the 6 groups were created by following Liu (2014) methodology, and these are:

- Grains
- Vegetables
- Fruits
- Meat, Poultry, and seafood
- Dairy
- Beans, eggs, and nuts

Different waves of CHNS use different food composition tables (FCT). Survey years 1989, 1991 and 1993 used 1981 CFCT; survey years 1997 and 2000 used 1991 CFCT; survey years 2004, 2006, 2009, 2011 and 2015 used 2002/04 FCT. While CFCT are similar, there are differences too. Therefore, it is not possible to include all waves without prior uniformization of the CFCT. I started my analysis by using FCT 2002/04. Additionally, I merged CFCT 1991 and 2002/04 to increase the sample. To do that, I translated 28 food groups from CFCT 1991 to 20 food groups from FCT 2002/02. While certain food groups were translated directly, such as "Cereals and cereal products", other food groups had to be split or merged. "Roots tubers and stems" food group from CFCT 1991 was split into "Tubers, starches and Products" as well as "Vegetable and Vegetable Products" of CFCT 2002/04. To do that, I analysed all individual foods from CFCT 1991, and determined to which food group of CFCT 2002/04 they should be placed. On the other hand, 3 food groups "Fishes", "Mollusks and Invertebrates" and "Crustaceans" from CFCT 1991 were associated to one food group "Fish, Shellfish, Mollusks" from FCT 2002/04. Two groups "Spice" and "Reptiles" from CFCT 1991 were not associated to any food group from FCT 2002/04, as their consumption was reported in a very few observations. Schematic presentation of food groups translation can be found in Annex 1 (Tables A.1.2 and A.1.3).

Famine:

The main independent variable of interest is famine. Since CHNS dataset does not contain explicit information on whether a respondent experienced famine or not, and if so, how severe the famine was, certain assumptions had to be made. To measure famine, I followed existing literature. As elaborated in the "Literature review" section there are two broad approaches. One is excess death rate in time of famine (EDR). The other is cohort size shrinkage, which presents the difference in cohort size during famine and non-famine years derived from 1990 Population Census. Furthermore, some authors chose to have the exposure to death rate as a binary variable – "Exposed Group" and "Not-exposed group" and they arbitrarily determine cut-off points (Li et al., 2010). Others chose to analyse EDR as a continuous variable (Meng et al., 2015; Xu et al., 2016).

In my work, I use a continuous variable Excess Death Rate (EDR) as a proxy for famine. This variable compares death rate (deaths per 1000) during the three years of famine (1959-1961) to death rate (deaths per 1000) three years prior to famine (1956-1958). As noted above, measuring EDR can be done in two ways.

The first measure is **ratio** between average value of death rate during the three years of famine, and average value of death rate during the three years prior to famine.

The second measure is **difference** between average death rate (deaths per 1000) during three years prior to famine and average death rate (deaths per 1000) during three years of famine. In the estimations I use both measures of EDR, so that I can verify if choice of measurement influence results. Provincial EDRs are presented in Annex 1, Table A.1.4.

$$EDR[\%] = \frac{Avg\ Famine\ (1959-1961)}{Avg\ Famine\ (1956-1958)}$$

or

$$EDR[diff] = Avg\ Famine(1959-61) - Avg\ Famine(1956-58)$$

In order for EDR to provide an accurate proxy for exposure to amine, 2 assumptions need to hold:

1. Mortality data published by the State Statistical Bureau (SSB) to be accurate,
2. Respondents did not move from one area to another – from area very affected by famine to area not affected at all, or in the opposite direction.

To minimize the measurement error which can emerge in case when a respondent moved from one location to another, in my estimations I used only those respondents with the same birth and interview location. This provides more confidence that a person has not changed the location. While an individual could have moved several times between birth and interview date, hukou system imposed by Chinese authorities limited mass migrations, especially from rural to urban areas. The mortality rate data from SSB come with a caveat – they are being measured on the provincial level, hence within-province differences in mortality rates cannot be captured in the current setup.

<u>Interaction terms:</u>

My estimations will include analysis of heterogenous effects of famine on dietary diversity score by gender, education and income categories. To that end, I will apply interaction terms EDR and income, as well as EDR and education. My hypothesis (*H1*) is that EDR and DDS are negatively correlated. Additionally, current literature argues that the relationship between income and DDS as well as education and DDS is a positive one. Therefore, to facilitate interpretation of interaction coefficients in my model, where one component of the interaction term is positively correlated with the dependent variable and another component is negatively correlated, I compute inverse EDR variable which I use only in estimations which contain the two interaction terms.

To compute *EDR_inv*, I first normalize *EDR* to a |0-1| scale:

$$NormalEDR_j = \frac{EDR_j - EDR_{min}}{EDR_{max} - EDR_{min}}$$

Then, in the second step I compute the inverse of this normalized variable:

$$EDR\_inv_j = 1 - NormalEDR_j$$

The resulting variable *EDR_inv* then captures the inverse of the excess death rate *EDR*, and as such, it is expected to have a positive association with dietary diversity score.

The interaction terms are as follows:

EDR_Inv x Income – presents interaction term between inverted EDR and log of household income

EDR_Inv x Education - presents interaction term between inverted EDR and years of education of head of household

EDR_Inv x Gender– presents interaction term between inverted EDR and gender of head of household, where the relationship between *Gender* and *DDS* proves to be positive one

<u>Age</u>

To estimate the effects of famine on dietary, nutrition and health outcomes, existing literature applies various methods to define age cohorts. As described in the "Literature review" (sub-section 2.1), some authors looked at birth years, and others of age of respondents during the famine. The first group of authors then relied on the respondents born before and after the famine, so that those born after the famine can act as a control group. In addition to that, some authors kept for the estimations only those born right after the famine (Almond et al., 2010), while other include those who were born up to 1975. The second group of authors were more interested in a respondents' age in the times of famine. Those analysed the following categories: unborn, infancy cohort, early childhood cohort and middle to late childhood cohort (Cheng & Smyth, 2021), prenatal, infant, early childhood, and adolescent periods (Herman et al., 2014), or childhood, adolescence, young adulthood, midlife and older (Ogle et al., 2013).

In my estimations, I use multi-step approach to capture age-based variations effects on DDS. I start with examination of impact of age on DDS. In this first step, I do not restrict sample to those born before or after certain date. In the following step, I include EDR and assume that the relationship between age and DDS is a liner one. In addition to that, I test whether this relationship is non-linear by introducing the quadratic term (Age^2). Then, I introduce interaction between Age and EDR to test for differentiated effects of age on those who experienced famine (Age x EDR). I also test for convex relationship in the interaction term (Age^2 x EDR). Additionally, I explore whether certain birth year cohorts were affected by the famine more than others. Those birth year cohorts are 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, and 1962. Table 1.1 below captures the difference in DDS between different birth-year cohorts over time.

Table 1. 1: Average Dietary Diversity Score (DDS) of different birth-year cohorts through 6 waves of CHNS

Panel B	Average Dietary Diversity Score				
	1935-39	1940-44	1945-49	1950-54	1955-59
1997	0.245	0.269	0.266	0.270	0.267
2000	0.254	0.276	0.283	0.282	0.277
2004	0.251	0.281	0.288	0.289	0.289
2006	0.273	0.281	0.291	0.315	0.306
2009	0.291	0.316	0.328	0.342	0.344
2011	0.306	0.329	0.348	0.370	0.368
Avg. Dietary Diversity Score	0.270	0.292	0.301	0.312	0.309

As we can see from Table 1.1 all five birth year cohorts increased DDS between 1997 and 2011, where each 0.05 value of DDS corresponds to one food group. Cohort 1935-39 DDS increased from 4.9 food groups to 6.1 food groups between 1997 and 2011. Cohort 1940-44 increased from 5.38 to 6.58, Cohort 1945-49 increased from 5.32 to 6.96, Cohort 1950-54 increased from 5.4 to 7.4 and finally, Cohort 1955-59 increased from 5.34 to 7.36 food groups between 1997 and 2011. In addition to this, those born between 1950 and 1954 on average have higher dietary diversity score than other birth-year cohorts. On average, they consumed 6.24 food groups in the given period, while other cohorts consumed between 5.4 and 6.18 food groups.

Finally, in my estimation where I test heterogenous effect of famine exposure on DDS by income, education and gender, I split sample into those born in the period 1954-1962 (famine cohort) and those born after 1962 (post-famine cohort). By doing this, I test whether the effects of famine exposure diminish over time, in generations born after famine.

Other variables

In addition to age, I use a vector of time-varying household level covariates: household income, household size, completed years of education, household head gender, household head nationality, household head marital status, keeping old home food habits and urban/rural areas. “Old home food habits” captures the household head preserved food habits from their parents’ place. Indirectly, it also captures generational spill-over effect. Summary statistics is presented in Table 1.2.

Exposure to famine is one of the main variables of interest, and I aimed to increase certainty on whether someone truly experienced famine, and if so, the severity of the effect. On one hand, CHNS does not contain explicit information on that, but it contains information where the interview took place and birthplace of a respondent. On the other hand, the dataset that I use for the mortality rate is on provincial level. To minimize risk of informants’ migration, and inaccurate data on whether they experienced famine, majority of authors who examined effects of the famine excluded urban areas completely, and there are two reasons for that. One is that famine was much more severe in rural areas than in urban areas, and there is broad consensus on this among authors who studied causes and effects of the famine. The second argument is that migrations in period during and after the famine were restricted, and that those who moved or who were relocated did this mostly from urban to rural areas, rather than in the opposite direction. If both of those arguments hold, that would have underestimated death rates in rural areas. While I do not disagree with the proposed arguments, I also think that this approach would not capture those who moved from one province to another within rural areas. Given that there is a large difference in EDR among provinces, this would result in under- or overestimation of exposure and severity to famine, and it is not possible to know whether it is under- or overestimate. Therefore, to limit noise driven by “movers”, I include only those respondents with the same birthplace and place of CHNS interview. In other words, this approach led to smaller, but more accurate sample.

When it comes to estimations, I combine different approaches, to perform sensitivity test. I estimate the specifications using both **EDRs** calculated as ratio and difference between three famine and pre-famine years. I denote them as *EDR* [%] and *EDR [diff].* Furthermore, I test consistency of results based on the number of **food groups**. Thus, I estimate the same specification using 20, 12 and 6 food groups. I use four **waves** separately as a cross section analysis, and I also pooled those 4 waves to increase power of my model. Those waves are 2004, 2006, 2009, 2011. In addition to that, I add two more waves, 1997 and 2001 to increase the sample size. The reason I did not include them initially is that those two waves used different China Food Composition Tables (CFCT) relative to other four waves. In data section, I explain the process how I merge information from the two CFCTs which allowed me to merge all 6 waves.

Summary statistics table of variables used in my models is presented below.

Table 1. 2: Summary Statistics of key variables
Dietary Diversity Score is a constructed variable based on 20 food groups – theoretical range of this value is from 0.1 to 1; Household total income is in 10,000 yuan; Education presents completed years of education of head of household; Household size presents number of persons living in the household; Age presents age of head of household in years, at the time of interview; Gender presents whether head of household is male (1) or female (0); Han nationality presents whether head of household belong to Han majority (1) or not (0); 'Old home' food habits presents whether the respondent keeps food consumption habits from parents place (1) or not (0); Marital status presents whether head of household is married (1) or not (0); Urban represents whether respondent lives in rural (1) or urban (0) area; EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years.

variable	mean	sd	min	p25	p50	p75	max	N
Dietary Diversity Score	0.29	0.11	0.10	0.20	0.30	0.35	0.80	5,770
HH total income	1.73	2.15	0	0.56	1.10	2.10	30.97	5,770
Education	6.46	3.36	0	5	6	9	17	5,770
HH size	4.22	1.40	1	3	4	5	11	5,770
Age	51	12	18	43	51	59	92	5,770
Gender	0.94	0.24	0	1	1	1	1	5,770
Han nationality	0.84	0.37	0	1	1	1	1	5,770
'Old home' food habit	0.96	0.19	0	1	1	1	1	5,770
Marital Status	0.95	0.22	0	1	1	1	1	5,770
Urban	0.03	0.17	0	0	0	0	1	5,770
EDR [%]	1.67	0.34	1.16	1.49	1.65	1.82	2.36	5,770
EDR [diff]	7.90	4.94	1.17	4.92	7.73	9.97	18.42	5,770

Summary statistics

In my estimations which include waves 1997-2011, sample size is 5770 (for waves 2004-2011 it is 3283). Furthermore, average *Dietary Diversity Score* is 0.29, which translates to 5.8 food groups. It should be noted though that the maximum value is 16 food groups (DDS=0.8). Average household monthly *Income* in the sample is 17300 yuan. Average completed *Years of education* of household head is 6.46 years, while their average Age is 51. 95% of household heads are married. Additionally, average *household size* is 4.2. 94% of the sample are *Males*. This share of male respondents results from restricting my sample to non-movers. 84% of my sample belong to *Han nationality*. 96% of the sample *preserve food habits* from their parents' place, and 3% of the sample is located in urban areas. Average value of *EDR ratio* is 1.67, where minimum value is 1.16 and maximum value is 2.36. Finally, Average value of *EDR difference* is 7.90, where minimum value is 1.17 and maximum value is 18.42.

Summary statistics presented above is to a large extent in line with other studies which use CHNS (Chen & Zhou, 2007; Fung & Ha, 2010; Luo et al., 2006). Discrepancies between summary statistics in this dissertation, and the ones used in other studies are caused by sample restriction, where I keep only "non-movers" while other authors do not impose that restriction.

Empirical strategy

In the previous section I describe the variables that I use in my analysis and elaborate on how and why I use different forms of the same variable in the estimations. In addition to that, I describe the rationale for restricting sample to non-movers, and for combining several waves. In the following section, I present my empirical strategy and explicitly present the models which I use in my analysis.

2.4.1 Determinants of dietary patterns

I begin my analysis by exploring the main determinants of dietary patterns, measured by Dietary Diversity Score (DDS). In particular, I first estimate the following cross-sectional specification for each of four waves in my sample: 2004, 2006, 2009 and 2011:

$$D_{h,j} = \sum_{v=1}^{V} \beta_v X_{v,h,j} + \varepsilon_{h,j} \tag{1}$$

Where $D_{h,j}$ denotes dietary diversity score of household *h* located in province *j*. $X_{v,h,j}$ is a vector which contains *v* household socio-economic variables, such as the logarithm of household total income, household size, completed years of education of the head of the household, gender, age and marital status of the head of the household, ethnicity, whether household is located in an urban or rural environment, and whether they maintain food habits from their previous home. As such, these tests will give us an indication of the main drivers of variation in dietary diversity scores.

2.4.2 Early life famine and dietary diversity score

The main goal of this *Section Two* is to assess the effect of early life exposure to famine on dietary diversity score later in life. To analyse this relation, I will augment Equation (1) to include a proxy for early life exposure to famine, as measured by the excess death rate (*EDR*) defined at the province level. In particular, I will estimate a series of cross-sectional regressions for each of the four waves (2004, 2006, 2009 and 2011):

$$D_{h,j} = \beta_1 EDR_j + \sum_{v=1}^{V} \beta_v X_{v,h,j} + \varepsilon_{h,j} \quad (2)$$

As a proxy for exposure to famine, I use EDR_j, which captures the excess death rate in province *j* in period 1959 - 1961.

Existing literature (Li & An, 2015; Li et al., 2010, 2011; Shi et al., 2013) has mostly focused on using select waves from the CHNS and other datasets in their analysis. A potential concern with this approach is that it gives rise to selection issues, in that drawing inference from a single study wave can have limitations when it comes to external validity of the obtained results. While in my approach above I use four different waves to conduct my cross-sectional analysis, a potential issue with the approach above is that it does not take into account the fact that certain time-specific factors prevalent in each wave could be driving the obtained results. To address this issue, I pool the four cross-sections together and estimate the following pooled cross-sectional regression:

$$D_{h,j,t} = \sum_{v=1}^{V} \beta_v X_{v,h,j,t} + \sum_{t=2004}^{2011} \theta_t + \sum_{j=1}^{12} \Pi_j + \varepsilon_{h,j,t} \quad (3)$$

Where $\sum_{t=2004}^{2011} \theta_t$ captures a vector of time-period (wave) dummies (i.e. for each wave 2004, 2006, 2009 and 2011), and $\sum_{j=1}^{12} \Pi_j$ captures a vector of province dummies. Note that in specifications that include vector of province dummy variables $\sum_{j=1}^{12} \Pi_j$, EDR coefficient cannot be estimated due to perfect collinearity with province dummies.

Motivated by Doan (2014), I examine if the relation between individual's age and their dietary diversity score is linear, by augmenting Equation (2) to include a quadratic term Age^2_h, which would capture any non-linear, that is convex structure between dietary diversity score and the age of the head of household *h*. In particular, I estimate a series of the following cross-sectional regressions for each of the four waves:

$$D_{h,j} = \beta_1 EDR_j + \beta_2 Age_{h,j} + \beta_3 Age_{h,j}^2 + \sum_{v=1}^{V} \beta_v X_{v,h,j} + \varepsilon_{h,j} \quad (4)$$

Where Age_h denotes the age of the head of household *h*, and EDR_j is defined as above.

As before, in addition to estimating four separate cross-sectional regressions for each wave, I will estimate the convexity of the relationship between the age of the head of the household and household dietary diversity score in a pooled cross-sectional setting. In particular, I will re-estimate Equation (3) to now include the quadratic age term:

$$D_{h,j,t} = \beta_1 \, Age_{h,j,t} + \beta_2 \, Age^2_{h,j,t} + \sum_{v=1}^{V} \beta_v \, X_{v,h,j,t} + \sum_{t=2004}^{2011} \theta_t + \sum_{j=1}^{12} \Pi_j + \varepsilon_{h,j,t} \qquad (5)$$

2.4.3 Early life famine and dietary patterns later in life: heterogenous effects by age

The test above allows us to analyse the relationship between household exposure to famine on dietary patterns later in life *on average*. To be able to distinguish between differential effects of famine exposure by age, and to explicitly test *Hypothesis H2*, I estimate the following specification:

$$D_{h,j} = \beta_1 \, EDR_j + \beta_2 \, Age_{h,j} + \beta_3 \, EDR_j \times Age_{h,j} + \sum_{v=1}^{V} \beta_v \, X_{v,h,j} + \varepsilon_{h,j} \qquad (6)$$

Where Age_h denotes age of the heads of households *h*, EDR_j is defined as above, and EDR_j *x* Age_h captures interaction between excess death rate and average age of the household head.

Driven by potential issues with cross-sectional analysis as discussed above, I repeat the analysis by pooling the four waves together and estimating a pooled cross-sectional specification:

$$D_{h,j,t} = \beta_1 \, Age_{h,j,t} + \beta_2 \, EDR_j \times Age_{h,j,t} + \sum_{v=1}^{V} \beta_v \, X_{v,h,j,t} + \sum_{t=2004}^{2011} \theta_t + \sum_{j=1}^{12} \Pi_j + \varepsilon_{h,j,t} \qquad (7)$$

The implicit assumption in specifications listed above is that there is a linear relationship between the average age of the household head and dietary patterns. In specification below, I relax this assumption and allow for a polynomial (convex) relationship between average age and dietary patterns. In particular, I first re-estimate Equation (6) by including a quadratic term Age^2_h for each of the four cross-sections:

$$D_{h,j} = \beta_1 \, EDR_j + \beta_2 \, Age_{h,j} + \beta_3 \, Age^2_{h,j} + \beta_4 \, EDR_j \times Age^2_{h,j} + \sum_{v=1}^{V} \beta_v \, X_{v,h,j} + \varepsilon_{h,j} \qquad (8)$$

I proceed in the same vein and re-estimate a pooled cross-sectional regression which now includes the interaction with the quadratic age term:

$$D_{h,j,t} = \beta_1 \, Age_{h,j,t} + \beta_2 \, Age^2_{h,j,t} + \beta_3 \, EDR_j \times Age^2_{hj,t} + \sum_{v=1}^{V} \beta_v \, X_{v,h,j,t} + \sum_{t=2004}^{2011} \theta_t + \sum_{j=1}^{12} \Pi_j + \varepsilon_{h,j,t} \qquad (9)$$

2.4.4 Early life famine and dietary patterns later in life: birth-cohort analysis

A potential concern with the above analysis is that it only captures an average effect and that it does not explicitly take into account the age (of the head) at the time of the famine and also at the time of measurement of dietary diversity. In this subsection I address this issue.

Based on existing population health literature (Dunn et al., 2019; Qin et al., 2020), I analyse whether the exposure to famine during the prenatal and early periods of childhood will exert larger effects than at other subsequent periods. As a result, the control group is made up of those individuals who were born after the famine (i.e., from 1963 to 1967)[1].

I first quantify the lasting effects of the famine on dietary diversity by estimating the following cross-sectional equations for each of the four waves:

$$D_{h,j} = \beta_1 EDR_j + \sum_{k=1954}^{1962} \gamma_k (EDR_j \times BirthYear_{i,k}) + BirthYear_k + \sum_{v=1}^{V} \beta_v X_{v,h,j} + \varepsilon_{h,j} \quad (11)$$

where $D_{h,j}$ is defined as above. $BirthYear_k$ are the year-cohort dummies, and EDR_j is the excess death rate in region j as defined before. $BirthYear_{i,k}$ is a dummy variable indicating whether individual *i* (head of household *h*) was born in the year *k*. Note that *k=1962* refers to a birth cohort in gestation in 1961 and born in 1962. The coefficient of the interaction between the excess death rate and birth cohort dummy variables measures the effect of the famine on dietary diversity score later in life. I expect that the magnitude of these estimated coefficients varies with birth cohorts. More specifically, I expect a larger impact of the famine on the relatively young birth cohorts, in particular those who were in gestation and early childhood during the famine.

Finally, I pool the four cross-sections together to estimate the same specification (11).

2.4.5 Early life famine and dietary patterns later in life: heterogenous effects by income

As elaborated above in the "*Methodology*" part, I use EDR inverse term in exploring the heterogenous effects of famine exposure on DDS by income and education. Here, I examine how differential exposure to early life famine affects dietary diversity score *across* the household income distribution.

I begin my analysis by assuming that the functional form of the relationship between dietary diversity and explanatory variables do not change over the examined period. That is, I estimate a series of cross-sectional OLS regressions for each surveyed year:

[1] My empirical results remain qualitatively and quantitatively similar if we include more birth cohorts born after 1967 as a control group.

$$D_{h,j} = \beta_1 \, EDR\_inv_j + \beta_2 \, Income_{h,j} + \beta_3 \, EDR\_inv_j \times Income_{h,j} + \sum_{v=1}^{V} \beta_v \, X_{v,h,j} + \varepsilon_{h,j} \qquad (12)$$

$$D_{h,j} = \beta_1 \, Income_{h,j} + \beta_2 \, EDR\_inv_j \times Income_{h,j} + \sum_{v=1}^{V} \beta_v \, X_{v,h,j} + \sum_{j=1}^{12} \Pi_j + \varepsilon_{h,j} \qquad (12.1)$$

Where $D_{h,j}$ denotes dietary diversity score of household *h* located in province *j*. $X_{v,h}$ is a vector of time-varying household socio-economic variables, such as household size, completed years of education of the head of the household, gender, age and marital status of the head of the household, ethnicity, whether household is located in an urban or rural environment, and whether they maintain food habits from their previous home. $Income_h$ is the logarithm of household *h* total income. Since vector of time-varying household and head of household socio-economic variables X_h includes household size, this gives the $Income_h$ term a *per-capita* interpretation. The coefficient of interest is β_3, as it captures the differential relation between income and dietary diversity for various levels of prior exposure to famine, as measured by *EDR_inv*. As described above in the "*Methodology*" part I create an inverse *EDR* variable *EDR_inv*. To compute *EDR_inv*, I first normalize *EDR* to a |0-1| scale:

$$NormalEDR_j = \frac{EDR_j - EDR_{min}}{EDR_{max} - EDR_{min}}$$

Then, in the second step I compute the inverse of this normalized variable:

$$EDR\_inv_j = 1 - NormalEDR_j$$

The resulting variable *EDR_inv* then captures the inverse of the excess death rate *EDR*, and as such, it is expected to have a positive association with dietary diversity score. $\sum_{j=1}^{12} \Pi_j$ denotes a vector of province dummy variables, which control for time-invariant unobserved heterogeneity at the province level that can be driving the results.

A potential drawback of the analysis presented above is that it assumes the functional form of the relationship between dietary diversity and explanatory variables does not change over time. In addition, it does not explicitly take into account time-specific unobserved heterogeneity (such as macro-economic factors that differ by survey year) that can be driving the results. To address this potential issue, I construct a pooled sample, by combing the four waves {2004, 2006, 2009 and 2011} and I estimate the following pooled cross-sectional regression:

$$D_{h,j,t} = \beta_1 \, EDR_{inv\,j} + \beta_2 \, Income_{h,j,t} + \beta_3 \, EDR_{inv\,j} \times Income_{h,j,t} + \sum_{v=1}^{V} \beta_v \, X_{v,h,j,t} + \sum_{t=2004}^{2011} \theta_t + \varepsilon_{h,j,t} \quad (13)$$

$$D_{h,j,t} = \beta_2 \, Income_{h,j,t} + \beta_3 \, EDR\_inv_j \times Income_{h,j,t} + \sum_{v=1}^{V} \beta_v \, X_{v,h,j,t} + \sum_{t=2004}^{2011} \theta_t + \sum_{j=1}^{12} \Pi_j + \varepsilon_{h,j,t} \quad (13.1)$$

Where $\sum_{t=2004}^{2011} \theta_t$ captures a vector of time-period (wave) dummies, and $\sum_{j=1}^{12} \Pi_j$ captures a vector of province dummies. Note that in specifications that include vector of province dummy variables $\sum_{1}^{12} \Pi$, β_1 cannot be estimated due to perfect collinearity with province dummies. Also, where I extend the sample and include waves 1997 and 2000, $\sum_{t=2004}^{2011} \theta_t$ becomes $\sum_{t=1997}^{2011} \theta_t$.

2.4.6 Early life famine and dietary patterns later in life: heterogenous effects by education level

Monotonic pattern of positive and increasing impacts of education on diet diversity is intuitive. Better educated people are likely to be more knowledgeable and/or more concerned about health and nutritional balance. Hence, it is expected that on average they make better informed food consumption decisions. The positive association between education and knowledge on one hand, and food choices on the other, has also been documented as described in "*Literature review*" section. Now, I try to disentangle this relation by analysing how prior exposure to famine affects the positive role of education on dietary patterns later in life.

In particular, I start by estimating a series of cross-sectional regressions for each survey year:

$$D_{h,j} = \beta_1 \, EDR\_inv_j + \beta_2 \, Education_{h,j} + \beta_3 \, EDR\_inv_j \times Education_{h,j} + \sum_{v=1}^{V} \beta_v \, X_{v,h,j} + \sum_{j=1}^{12} \Pi_j + \varepsilon_{h,j} \quad (14)$$

As before, the implicit assumption here is that the functional form of the relationship between dietary diversity and explanatory variables does not change over the examined period. $D_{h,j}$ and EDR_inv_j are defined as above. Specifications also contain a vector v of time-varying household socio-economic variables X_h defined as above, as well as a vector of province dummy variables $\sum_{j=1}^{12} \Pi_j$. As before, note that the inclusion of $\sum_{j=1}^{12} \Pi_j$ implies that β_1 will not be estimated in the regression, since EDR_inv_j is perfectly colinear with $\sum_{j=1}^{12} \Pi_j$. $Education_h$ denotes years of completed education for the head of household *h.*

The coefficient of interest is β_3, as it captures the differential relation between education and dietary diversity for various levels of prior exposure to famine, as measured by *EDR_inv*.

As above, I then also conduct a pooled cross-sectional regression of the following form:

$$D_{h,j,t} = \beta_1 \, EDR\_inv_j + \beta_2 \, Education_{h,j,t} + \beta_3 \, EDR\_inv_j \times Education_{h,j,t} + \sum_{v=1}^{V} \beta_v \, X_{v,h,j,t} + \sum_{t=2004}^{2011} \theta_t + \varepsilon_{h,j,t} \quad (15)$$

$$D_{h,j,t} = \beta_2 \, Education_{h,j,t} + \beta_3 \, EDR\_inv_j \times Education_{h,j,t} + \sum_{v=1}^{V} \beta_v \, X_{v,h,j,t} + \sum_{t=2004}^{2011} \theta_t + \sum_{j=1}^{12} \Pi_j + \varepsilon_{h,j,t} \quad (15.1)$$

Where as above, $\sum_{t=2004}^{2011} \theta_t$ captures a vector of time-period (wave) dummies, and $\sum_{j=1}^{12} \Pi_j$ captures a vector of province dummies. Note that in specifications that include vector of province dummy variables $\sum_{j=1}^{12} \Pi_j$, β_1 cannot be estimated due to perfect collinearity with province dummies.

2.4.7 Early life famine and dietary patterns later in life: heterogenous effects by gender

Existing literature indicates that households whose heads are male tend to have higher dietary diversity. Here, I further explore this relationship by analysing exposure to famine and its effects on household DDS, differentiated by gender.

In particular, I start by estimating a series of cross-sectional regressions for each survey year:

$$D_{h,j} = \beta_1 \, EDR\_inv_j + \beta_2 \, Gender_{h,j} + \beta_3 \, EDR\_inv_j \times Gender_{h,j} + \sum_{v=1}^{V} \beta_v \, X_{v,h,j} + \varepsilon_{h,j} \quad (16)$$

$D_{h,j}$ and EDR_inv_j are defined as above. Specifications also contain a vector of time-varying household socio-economic variables X_h defined as above. $Gender_h$ denotes gender of the head of household *h:* $Gender_h = 1$ if the head of the household *h* is male, and $Gender_h = 0$ otherwise.

The coefficient of interest is β_3, as it captures the relation between head of household gender and dietary diversity for various levels of prior exposure to famine, as measured by *EDR_inv*. Put differently, it measures whether and how differently did famine affect male vs female headed households in terms of their dietary diversity later in life.

As above, I then also conduct a pooled cross-sectional regression of the following form:

$$D_{h,j,t} = \beta_1 \, EDR\_inv_j + \beta_2 \, Gender_{h,j} + \beta_3 \, EDR\_inv_j \times Gender_{h,j} + \sum_{v=1}^{V} \beta_v \, X_{v,h,j,t} + \sum_{t=2004}^{2011} \theta_t \; + \varepsilon_{h,j,t} \quad (17)$$

Where as above, $\sum_{t=2004}^{2011} \theta_t$ captures a vector of time-period (wave) dummies. Note that in these specifications I cannot include province dummies, $\sum_{j=1}^{12} \Pi_j$, since they are perfectly collinear with both EDR_inv_j and $EDR\_inv_j \times Gender_h$.

2.5 Results

The existing literature models dietary patterns using various socio-economic variables. Education, income, ethnicity, urban residence, marital status, nutritional and dietary knowledge have found to be positively associated with dietary diversify score, while smoking and alcohol consumption have been negatively associated to DDS (Hou et al., 2021; Wang et al., 2021; Zhang et al., 2020; Zhang & Zhao, 2021; Zhang et al., 2017; Zhao et al., 2020; Zhao et al., 2017; Zhong et al., 2018). At the same time, evidence on the impact of age and gender on dietary diversify score has been mixed. Given the potentially long-lasting effects of exposure to extreme life changing events, such as wars, natural disasters and famine, little is yet known about the effects of these events on dietary diversity patterns later in life. In this Section I discuss the results of estimating these relationships using the CHNS data set.

2.5.1 Dietary patterns – multivariate analysis

I begin my analysis by estimating Equation (1) using the CHNS dataset covering 4 survey waves (2004, 2006, 2009 and 2011). *Table 1.3* shows the results of cross-sectional OLS regressions for each wave as well as pooled regressions, where the dependent variable $D_{h,j}$ denotes average dietary diversity score of household *h* located in province *j.* $D_{h,j}$ is computed using the 20 food groups from the Food Composition Tables described in "*Methodology and empirical strategy"* section.[2] Columns 1-4 (4 separate waves cross sectional) as well as column 9 (pooled regression waves 1997-2011)

[2] In Section 2.5.4 "Alternative measures of DDS", I re-estimate and discuss our results when $D_{h,j}$ is computed using the 6 and 12 food groups from the Food Composition Tables.

contain *Age* (head of household) variable, while columns 5-8 (4 separate waves cross sectional) as well as column 10 (pooled regression waves 1997-2011) in addition to *Age* contain *Age²* variable.

As we can see from *Table 1.3*, the main drivers of variation in DDS are income, education, Han nationality marital status and urban areas which are positively correlated with DDS The estimated coefficient on household size is negative.

Table 1. 3: Determinants of Dietary Diversity Score, Cross sectional regressions, pooled regressions (1997-2011)
*Variables are defined as described in Table 1.2. Dependent variable Dietary Diversity Score is calculated using 20 food groups. Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; pooled regressions include wave dummies.*

	Dietary Diversity Score									
	2004	2006	2009	2011	2004	2006	2009	2011	Pooled 1997-2011	Pooled 1997-2011
log (HH total income)	0.079***	0.068***	0.045***	0.066***	0.078***	0.068***	0.045***	0.066***	0.062***	0.061***
	[9.774]	[7.490]	[6.035]	[8.899]	[9.666]	[7.419]	[6.008]	[8.859]	[19.244]	[18.733]
Education	0.002	0.004***	0.003**	0.002	0.002	0.004***	0.003**	0.001	0.003***	0.003***
	[1.638]	[3.338]	[2.275]	[1.131]	[1.546]	[3.261]	[2.296]	[0.981]	[5.782]	[5.724]
HH size	-0.011***	-0.012***	-0.008**	-0.005	-0.012***	-0.014***	-0.009***	-0.006*	-0.008***	-0.009***
	[-4.586]	[-4.050]	[-2.548]	[-1.463]	[-4.875]	[-4.599]	[-2.779]	[-1.737]	[-8.224]	[-8.772]
Gender	0.002	-0.003	0.026	0.010	0.002	-0.002	0.026	0.012	-0.003	-0.003
	[0.153]	[-0.193]	[1.577]	[0.675]	[0.206]	[-0.121]	[1.589]	[0.774]	[-0.524]	[-0.496]
Age	0.015	0.105**	0.016	-0.109*	0.591**	1.055**	0.692	0.737	0.028*	0.395***
	[0.428]	[2.277]	[0.288]	[-1.866]	[2.166]	[2.560]	[1.424]	[1.240]	[1.813]	[3.816]
Han nationality	0.015**	0.038***	0.017*	-0.002	0.016**	0.039***	0.018**	-0.001	0.009***	0.010***
	[2.042]	[4.792]	[1.937]	[-0.198]	[2.199]	[4.899]	[2.048]	[-0.108]	[2.639]	[2.838]
Old home food habits	0.003	0.018	-0.025	0.029	0.003	0.018	-0.025	0.030	0.001	0.001
	[0.146]	[0.997]	[-1.177]	[1.116]	[0.134]	[1.002]	[-1.150]	[1.157]	[0.179]	[0.169]
Marital Status	0.065***	0.048***	0.018	0.031	0.063***	0.047***	0.014	0.030	0.037***	0.036***
	[5.178]	[2.658]	[0.735]	[1.333]	[4.922]	[2.644]	[0.575]	[1.250]	[5.832]	[5.742]
Urban	0.033***	0.014	0.041**	0.028	0.033***	0.013	0.041**	0.029	0.026***	0.026***
	[2.982]	[0.839]	[2.055]	[1.396]	[2.985]	[0.782]	[2.055]	[1.441]	[5.225]	[5.206]
Age^2					-0.515**	-0.825**	-0.565	-0.687		-0.333***
					[-2.186]	[-2.355]	[-1.399]	[-1.418]		[-3.546]
Wave Dummies									Y	Y
Observations	1,001	832	764	686	1,001	832	764	686	5,748	5,748
R-squared	0.144	0.145	0.098	0.16	0.147	0.151	0.1	0.163	0.174	0.176

The estimated coefficient on *log(HH total income)* is positive and significant at 1% level in all waves, in both cross sectional and pooled regressions. It ranges from 0.045 to 0.079, suggesting that households with higher total income tend to have higher dietary diversity score. Furthermore, the coefficients suggests that 1% in increase of income corresponds to increase of food groups intake between 0.009 to 0.02. While this appears to be a very small change, it should be noted that as per the summary statistics table, the mean household income is 17,300 yuan and the standard deviation is 21,500yuan. Given that in all specifications I explicitly control for household size (*HH size*), this gives the above coefficient a "per capita" interpretation.

The estimated coefficients on *Completed Years Education (Head)* are also positive across all waves. It is statistically significant at 1% level in 2006, and in pooled regressions, and at 5% level in 2009, whereas in 2004 and in 2011 it is not statistically different from 0. In 2006 and 2009 it ranges from 0.003 to 0.004, suggesting that households with higher average number of years in education tend to have higher dietary diversity score.

The estimated coefficients on Household size (*HH size*) are negative in all four waves, and statistically significant at 1% level in 2004 and in 2006, and in pooled regressions, suggesting that households with higher number of household members tend to have lower dietary diversity score.

The estimated coefficient on *Age* (of the head of household) is positive in all four waves, with the exception of 2011, and statistically significant in 2004 and in 2006. Interestingly, the estimated coefficients in cross sectional regressions increased in magnitude after introducing the quadratic term Age^{2}_{head}. As for the quadratic term, all coefficients are negative, yet waves 2004 and 2006 are significant at 5%. While the estimated coefficients on Age_{head}^{2} are not precisely estimated for 2009 and 2011 waves, they however suggest the presence of a convex relationship between head of household's age, and their dietary diversity score later in life.

Marital status of the head of the household enters with a positive sign in all four waves: the estimated coefficient is statistically significant at 1% level in 2004 and in 2006, and ranges from 0.047 to 0.065 in cross sectional and between 0.036-0.037 in pooled regressions, suggesting that households in which the head of the household is married tend to have higher dietary diversity scores.

When looking at the nationality of the head of the household, the estimated coefficient on *Han nationality dummy* variable is positive and significant in waves 2004, 2006 and 2009 and negative but statistically not significant in 2011. In both pooled regression the coefficient is positive and significant at 1%.

The estimated coefficient on the *Urban* dummy variable is positive and significant at 1% level in 2004 and in 2009, ranging from 0.033 to 0.041, suggesting that household located in urban provinces tend to have higher dietary diversity scores. Pooled regressions confirm this as the coefficient is positive and significant at 1%, with value 0.026.

Estimated coefficients on *Old Home Food Habits* are positive, but statistically not significant at conventional levels.

Evidence on the *Gender of the Head* is more mixed, but statistically indistinguishable from zero.

2.5.1.1 Pooled regressions

I start my analysis by applying pooled regressions on four waves {2004, 2006, 2009 and 2011}, and then I proceed with cross-sectional regressions for the purpose of robustness check. The reason I apply the pooled regressions first are multiple. A potential concern with the cross-sectional analysis presented above is that there could have been macro-economic factors present in each of the waves that are driving my analysis, but which could not have been accounted for due to the cross-sectional empirical design. Furthermore, the cross-sectional analysis could also be biased due to potential selection issues, in that using individual waves to conduct cross-sectional analysis provides limited data samples on which regressions are run, and thus reducing the statistical power.

I start with estimating Equation (3) and Equation (5):

Results of this estimation are shown in *Table 1.4*. Columns (1) and (2) show the results of estimating Equation (3), with *EDR* [%] and *EDR [diff]* used as proxies for prior exposure to famine, respectively. The estimated coefficients on *EDR* [%] is negative and statistically significant at 1% level (Column (1)). In the same vein, the estimated coefficient on *EDR [diff]* is negative and statistically significant at 1% level (Column (2)). As for the the economic magnitudes, the negative coefficient of -0.081 translates to 1.62 food groups, hence for each unit increase in EDR [%], the number of food groups consumed reduces by 1.62. Similarly, for each unit increase in EDR [diff], the number of food groups consumed reduces by 0.12.

Columns (3) and (4) show the results of estimating Equation (5), with *EDR [%]* and *EDR [diff]* used as proxies for prior exposure to famine, respectively. The estimated coefficient on *EDR* [%] is negative and statistically significant at 1% level (Column (3)). In the same vein, the estimated coefficient on *EDR [diff]* is negative and statistically significant at 1% level (Column (4)). The economic magnitudes of the estimates are similar to those reported in Columns (1) and (2). Importantly, the estimated coefficients on *Age* are positive and significant at 1% level in both Column 3 and Column 4. At the same time, the estimated coefficient on the quadratic term Age_h^2 is negative and statistically significant at 1% level in both columns, suggesting a significant convex relation between household head's age at their dietary diversity score later in life.

Finally, AIC and BIC suggest that including Age^2 in model does improve its relative goodness-of-fit.

Table 1. 4: Effects of famine (EDR [%/diff]) on Dietary Diversity Score; Pooled regressions (2004-2011)

Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using 20 food groups
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include wave dummies.*

	(1)	(2)	(3)	(4)
	Dietary Diversity Score			
log(HH total income)	0.060***	0.060***	0.060***	0.059***
	[15.281]	[15.235]	[15.127]	[15.073]
Education	0.002***	0.002***	0.002***	0.002***
	[2.937]	[2.720]	[2.871]	[2.649]
HH size	-0.006***	-0.006***	-0.007***	-0.007***
	[-4.440]	[-3.986]	[-4.933]	[-4.507]
EDR [%]	-0.081***		-0.078***	
	[-11.152]		[-10.767]	
Age	0.008	0.011	0.529***	0.540***
	[0.338]	[0.484]	[2.680]	[2.751]
Gender	0.000	0.000	0.001	0.001
	[0.051]	[0.009]	[0.142]	[0.095]
Han nationality	-0.017***	-0.017***	-0.015***	-0.015***
	[-3.333]	[-3.404]	[-2.927]	[-3.046]
Old home food habits	0.009	0.011	0.009	0.011
	[0.820]	[1.103]	[0.817]	[1.098]
Marital Status	0.037***	0.035***	0.035***	0.034***
	[3.935]	[3.797]	[3.778]	[3.639]
Urban	0.032***	0.033***	0.032***	0.033***
	[4.048]	[4.204]	[4.035]	[4.191]
EDR [diff]		-0.006***		-0.006***
		[-12.460]		[-12.160]
Age^2			-0.447***	-0.454***
			[-2.650]	[-2.702]
Wave Dummies	Y	Y	Y	Y
Observations	3,283	3,283	3,283	3,283
R-squared	0.192	0.199	0.194	0.201
AIC	-5,477	-5,504	-5,482	-5,509
BIC	-5,392	-5,418	-5,391	-5,418

While the section below analyzes the coefficients from the cross-sectional regression in detail, I first compare the coefficients, their magnitude and statistical significance between pooled regressions and cross-sectional regressions. My main variable of interest **- EDR**, is very similar in magnitude and statistical significance, and this applies to both measures of EDR. Similarly, the magnitude and statistical significance of **Income** is similar in pooled regressions and cross-sectional regressions. When it comes to **Education**, **Size of the Household**, and **Marital status**, magnitude is similar in pooled regressions and cross-sectional regressions, while the statistical significance is consistently higher in the pooled regressions. As for **Gender**, while cross-sectional regressions reveal that there are certain variations among the waves, on average, the results from those regressions confirmed the results from the pooled regressions in magnitude, while the statistical significance is slightly lower. Finally, the coefficient which depicts convex nature of Age, follows the same pattern as Gender – pooled regressions have higher statistical significance and similar magnitude comparing to cross-sectional regressions.

2.5.1.2. Cross-sectional regressions

As a robustness check, and also to verify whether certain waves are outliers, which can add to bias to my estimations, I perform cross sectional analysis of the four waves {2004, 2006, 2009 and 2011}. I start my analysis by estimating Equation (2). *Table A. 1.5* shows the results of cross-sectional OLS regressions for each wave. Dependent variable $D_{h,j}$ has been described above.

Columns (1)-(4) show the results of estimating Equation (2) for each wave {2004, 2006, 2009 and 2011}, where the vector of independent variables includes: natural logarithm of household total income, household size, (household head's) completed years of education, age, gender and marital status of the head of the household, (household head's) nationality, whether household is located in an urban or rural environment, and whether they maintain food habits from their previous home. All specifications include robust standard errors that are adjusted for heteroscedasticity[3].

The main independent variable of interest *EDR [%]* captures the excess death rate in period 1959-1961 in province *j*, where the household is located. These variables as such serve as a proxy for the household's previous exposure to famine (Xu et al., 2016). As we can see from *Table 1.4*, the estimated coefficients on excess death rate *EDR* [%] are negative and significant at 1% level in all four waves, ranging from 1.28 food groups (-0.064 in 2004) to 1.86 food groups (-0.093 in 2009). These results suggest that households that were more exposed to Great China Famine or located in provinces with more severe famine as proxied by the excess death rates in period 1959-1961, tend to have lower average dietary diversity scores decades after the event.

The estimated coefficient on *log (HH total income)* is positive and significant at 1% level in all waves, and ranges from 0.043 to 0.076, suggesting that households with higher total income tend to have higher dietary diversity score. Given that in all specifications we explicitly control for household size (*HH size*), this gives the above coefficient a "per capita" interpretation.

The estimated coefficients on *Completed Years Education (Head)* are also positive across all waves. It is statistically significant at 1% level in 2006, and at 10% level in 2009, whereas in 2004 and in 2011 it is not statistically different from 0. In 2006 and 2009 it ranges from 0.002 to 0.003, suggesting that households with higher average number of years in education tend to have higher dietary diversity score.

The estimated coefficients on Household size (*HH size*) are negative in all four waves, and statistically significant at 1% level in 2004 and in 2006 (-0.009 in both waves), suggesting that households with

[3] Results remain quantitively similar when clustering standard errors by province.

higher number of household members tend to have lower dietary diversity score. While the magnitudes are similar to the ones obtained in the pooled regressions, the statistical significance is lower.

Marital status of the head of the household enters with a positive sign in all four waves: the estimated coefficient is statistically significant at 1% level in 2004 and in 2006, and ranges from 0.037 to 0.060, suggesting that households in which the head of the household is married tend to have higher dietary diversity scores. Similarly to the previous two variables, the coefficients are comparable to those from the polled regression, yet he statistical significance is lower.

When looking at the nationality of the head of the household, the estimated coefficient on *Han nationality dummy* variable is negative and significant at 1% level (5% level) in 2011 (2009). In 2004 and in 2006 it cannot be statistically distinguished from 0. The coefficients range from 0.002 to -0.038, offering mixed evidence between Han nationality and dietary diversity score when using cross-sectional analysis.

The estimated coefficient on the *Urban* dummy variable is positive and significant at 1% level in 2004 and in 2009, ranging from 0.034 to 0.049, suggesting that household located in urban provinces tend to have higher dietary diversity scores. Estimated coefficients on *Old Home Food Habits* are positive, but statistically not significant at conventional levels. Given that 96% of the sample stated that they preserved food habits from their parents' place, and that the famine was one of the most disastrous in recorded history of famines, it is surprising that this coefficient has shown no statistical significance. Evidence on the *Gender of the Head* is more mixed, but statistically indistinguishable from zero. While cross-sectional regressions reveal that there are certain variations among the waves, on average, the results from those regressions confirmed the results from the pooled regressions in magnitude, while the statistical significance is slightly lower.

In *Table A. 1.6*, I re-estimate Equation (2) by using *EDR [diff]* as our main independent variable of interest. *EDR [diff]* is computed by subtracting average mortality in period 1956-1958 from average mortality in period 1959-1961. The estimated coefficient on *EDR [diff]* is negative and significant at 1% level in all four waves, ranging from -0.005 (in 2004 and 2011) to -0.007 (in 2009). This translates to 0.1 food group in 2004 and 2011, and 0.14 food groups in 2009. Estimated coefficients on all other covariates remain qualitatively and quantitatively similar to those reported in *Table A. 1.5*. These results proved further evidence that households that were more exposed to Great China Famine, or located in provinces with more severe famine, irrespective of the method used to compute the excess death rates, tend to have lower average dietary diversity scores later on in life. Having said that, the magnitude of the effect of EDR on DDS depends on the way I measure EDR.

2.5.2 Convexity in age

2.5.2.1 Cross-sectional analysis

In this part, I relax the assumption that the relationship between age and dietary diversity score is linear and allow for convexity in this relationship. In particular, I introduce a quadratic age term Age_{head}^2 into my multivariate analysis, as shown in Equation (4). *Table A. 1.7* shows the results of estimating Equation (4), using the *EDR* [%] as the proxy for exposure to famine, whereas *Table 1.7* shows the results using *EDR [diff]* as the proxy for exposure to famine.

The estimated coefficient on *EDR [%]* in *Table A.1.7.* is still negative and significant at the 1% level in all four cross-sectional regressions, ranging from -0.062 in the 2004 wave, to -0.091 in the 2009 wave. This translates to 1.24 food groups in 2004 and 1.82 food groups in 2009 wave. The estimated coefficient on *Age* (of the head of household) is positive in all four waves, and statistically significant at 10% level in 2004 and in 2006. Interestingly, the estimated coefficient on the quadratic term, Age_h^2 is negative in all four waves, however, it is only statistically significant at 10% level for the 2004 wave. While the estimated coefficients on Age_h^2 are not precisely estimated for 2006, 2009 and 2011 waves, they however suggest the presence of a convex (concave) relationship between head of household's age, and their dietary diversity score later in life. The estimated coefficients on other control variables are very similar in sign, magnitude and statistical significance as those reported in *Table 1.4*.

In *Table A.1.8* we re-estimate Equation 4, but now we use *EDR [diff]* as a proxy for exposure to famine. The estimated coefficient on *EDR [diff]* continues to be negative, and statistically significant at 1% in all four waves. The estimated coefficients are similar to those reported in *Table 1.5*, and range from -0.004 in 2004, to -0.007 in 2009. This translates to 0.08 food groups in 2004 and 0.14 food groups in 2009. The estimated coefficient on *Age* (of the head of household) is positive in all four waves, and statistically significant at 10% level in 2004 and in 2006. Interestingly, the estimated coefficient on the quadratic term, Age_h^2 is negative in all four waves, however, it is only statistically significant at 10% level for the 2004 wave. The estimated coefficients on other control variables are very similar in sign, magnitude and statistical significance as those reported in *Table 1.5*.

Taken together with the evidence presented in *Table A.1.7*, this suggests that the relationship between head of household age and household dietary diversity score is not linear. I will explore this more formally in Section 2.5.3 “Heterogenous effects of famine by age”. Also, when comparing the coefficients from pooled and cross-sectional regressions, we can see that the magnitude of the coefficients are either similar or slightly lower in cross-sectional regressions, yet the statistical significance is consistently higher in the pooled regressions. Given the sample size and the statistical power that polled regressions hold, these findings do not come as a surprise.

2.5.3 Heterogenous effects of famine by age

Evidence presented in Section 2.5.1 suggests that households that were in the past more exposed to famine or located in provinces with more severe famine, on average tend to have lower dietary diversity score later on in life. A potential concern with the above analysis is that it only captures an average effect and that it does not explicitly take into account the age (of the head) at the time of the famine and also at the time of measurement of dietary diversity score. In this Section I address this issue.

2.5.3.1 Pooled cross-sectional results

Given the above-mentioned concerns with cross-sectional analysis, I pool the four cross-sections together and estimate Equation (7) and Equation (9). Results of these pooled regressions are shown in *Table 1.5*.

Table 1. 5: Effects of interaction famine of (EDR [%/diff]) and Age/Age² on Dietary Diversity Score; Pooled regressions (2004-2011)

Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using 20 food groups
Robust standard errors are adjusted for heteroscedasticity; *significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively*; *t-statistics in brackets*. All regressions include wave dummies, while regressions (5)-(8) include both wave and province dummies.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Dietary Diversity Score				
log (HH total income)	0.060***	0.060***	0.060***	0.059***	0.050***	0.050***	0.050***	0.050***
	[15.277]	[15.246]	[15.096]	[15.064]	[13.108]	[13.147]	[12.876]	[12.927]
Education	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	[2.935]	[2.748]	[2.854]	[2.670]	[2.867]	[2.917]	[2.786]	[2.847]
HH size	-0.006***	-0.006***	-0.007***	-0.007***	-0.000	-0.000	-0.001	-0.001
	[-4.471]	[-4.029]	[-5.036]	[-4.640]	[-0.230]	[-0.257]	[-0.917]	[-0.944]
EDR [%]	-0.136***		-0.119***					
	[-3.790]		[-6.114]					
Age	-0.156	-0.051	0.599***	0.625***	-0.340***	-0.121***	0.564***	0.560***
	[-1.382]	[-1.152]	[2.999]	[3.141]	[-3.046]	[-2.782]	[2.886]	[2.861]
EDR [%] x Age	0.094				0.178***			
	[1.525]				[2.910]			
Gender	0.001	0.001	0.002	0.002	-0.006	-0.006	-0.005	-0.005
	[0.127]	[0.094]	[0.256]	[0.221]	[-0.857]	[-0.878]	[-0.722]	[-0.741]
Han nationality	-0.017***	-0.016***	-0.014***	-0.014***	0.003	0.003	0.006	0.005
	[-3.254]	[-3.263]	[-2.799]	[-2.827]	[0.642]	[0.602]	[1.041]	[1.001]
Old home food habits	0.008	0.011	0.008	0.011	0.019*	0.019*	0.019*	0.019*
	[0.805]	[1.084]	[0.801]	[1.079]	[1.921]	[1.904]	[1.918]	[1.896]
Marital Status	0.038***	0.036***	0.037***	0.036***	0.035***	0.035***	0.034***	0.034***
	[4.017]	[3.905]	[3.889]	[3.781]	[3.833]	[3.820]	[3.708]	[3.701]
Urban	0.033***	0.034***	0.033***	0.034***	0.055***	0.055***	0.056***	0.055***
	[4.144]	[4.322]	[4.186]	[4.373]	[6.971]	[6.948]	[7.005]	[6.984]
EDR [diff]		-0.010***		-0.009***				
		[-4.302]		[-6.904]				
EDR [diff] x Age		0.007*				0.010***		
		[1.784]				[2.638]		
Age²			-0.710***	-0.603***			-0.830***	-0.604***
			[-3.365]	[-3.359]			[-3.987]	[-3.417]
EDR [%] x Age²			0.116**				0.183***	
			[2.145]				[3.404]	
EDR [diff] x Age²				0.009**				0.011***
				[2.460]				[3.175]
Wave Dummies	Y	Y	Y	Y	Y	Y	Y	Y
Province Dummies					Y	Y	Y	Y
Observations	3,283	3,283	3,283	3,283	3,283	3,283	3,283	3,283
R-squared	0.193	0.200	0.195	0.202	0.272	0.272	0.274	0.274
AIC	-5,477	-5,504	-5,485	-5,513	-5,804	-5,803	-5,812	-5,811
BIC	-5,386	-5,413	-5,387	-5,415	-5,676	-5,675	-5,678	-5,677

Columns 1-4 show the results of estimating Equation (9) using wave dummies, and columns 5-8 show the results of estimating Equation (9) with wave and province dummies. Odd columns show results with *EDR* [%] as a proxy for exposure to famine, while even columns use *EDR [diff].*

As we can see from Column 1, the estimated coefficient on *EDR* [%] is negative and statistically significant at 1% level, confirming my prior results. The coefficient -0.136 corresponds to 2.72 food groups consumed less for each unit of EDR increase. Interestingly, the estimated coefficient on the interaction term *EDR x Age (Head)* is positive, yet imprecisely estimated. As Column 2 shows, the estimated coefficient on *EDR [diff]* is negative and statistically significant at 1% level, and of larger magnitude to those presented in *Table 1.8*. Coefficient value -0.010 corresponds to 0.2 food groups. Interestingly, the estimated coefficient on the interaction term *EDR [diff] x Age (Head)* is positive, and significant at 10% level (0.007).

In Column 3, I add the interaction between exposure to famine and the quadratic age term, *EDR x* Age_h^2. The estimated coefficient on this interaction term is positive and statistically significant at 5% level (0.116), which corresponds to 0.32 food groups. In Column 4 I instead use the interaction between exposure to famine and the quadratic age term, *EDR [diff] x* Age_h^2. The estimated coefficient on this interaction term is positive and statistically significant at 5% level (0.009).

Columns 5-8 show the results of my most saturated pooled regressions, which include wave and province dummies. Note that in these specifications, I do not include *EDR* [%] (*EDR [diff],* respectively), since they are time-invariant at the province level, and due to inclusion of province fixed effects they cannot be estimated. As we can see from Column 5, the estimated coefficient on the on the interaction term *EDR [%] x Age (Head)* is positive, and statistically significant at 1% level (0.178, or 3.56 food groups). Similarly, in Column 6, the estimated coefficient on the on the interaction term *EDR [diff] x Age (Head)* is positive, and statistically significant at 1% level (0.010, or 0.2 food groups), further confirming my prior results.

In Column 7, I add the interaction between exposure to famine and the quadratic age term, *EDR x* Age^2 *(Head).* The estimated coefficient on this interaction term is positive and statistically significant at 1% level (0.183), and larger in magnitude than the one reported in Column 3, where I did not control for province dummies. In Column 8 I instead use the interaction between exposure to famine and the quadratic age term, *EDR [diff] x* Age^2 *(Head).* The estimated coefficient on this interaction term is positive and statistically significant at 1% level (0.011).

Taken together, evidence presented in this section suggests that the negative effects of exposure to famine have not been experienced equally across all age cohorts. Individuals who were older at the time of interview the negative effect of famine on DDS is attenuated.

In addition to the coefficients from the previous pooled regression, I add the interaction term between EDR and Age in this one. Comparing to the coefficients in the pooled regression, the coefficient in cross-sectional are higher in magnitude, yet with lower statistical significance, and that findings is common for EDR (%) and EDR (diff), but also for the interaction between EDR and quadratic Age term.

2.5.3.2 Cross-sectional results

As a robustness check, and also to verify whether certain waves are outliers, which can add to bias to my estimations, I perform cross sectional analysis of the four waves {2004, 2006, 2009 and 2011}. I first estimate Equation (6), that explicitly models the interaction between exposure to famine (*EDR*) and the *Age* of the head of household: *EDR [%] x* Age_h (*EDR [diff] x* Age_h*,* respectively)*.* Results of this cross-sectional estimation for each four waves are shown in *Table 1.9* (*Table 1.10*, respectively).

The estimated coefficient on *EDR* [%] in *Table 1.9* is negative and statistically significant at 1% level in all four waves, except of 2009, where is positive but not precisely estimated. The estimated coefficient on *Age* is negative in 2004, 2006 and 2011, and positive in 2009, while only being statistically significant at 5% level in 2011 wave. Interestingly, the estimated coefficient on the interaction term *EDR x Age (Head)* is mostly positive (with the exception of 2009), and statistically significant at 10% level in 2004 and 2011. 2004 and 2011 coefficients 0.166 and 0.3 respectively suggest that the two variables, EDR and Age, have compounding negative effect on DDS.

Turning our attention to *Table 1.10*, the estimated coefficient on *EDR[diff]* in *Table 1.10* is negative and statistically significant at 1% level in 2004, 5% level in 2006 and in 2011. The coefficient magnitude in 2004, 2006 and 2011 translate to 0.22, 0.2 and 0.32 food groups respectively. The estimated coefficient on *Age* is negative in 2004, and 2011, and positive in 2006 and in 2009, while only being statistically significant at 5% level in 2011 wave. Interestingly, the estimated coefficient on the interaction term *EDR[diff] x Age (Head)* is mostly positive (with the exception of 2009), and statistically significant at 10% level in 2004 and 2011, confirming our results from *Table 1.9*. Similarly to previous estimations, coefficients obtained by using EDR [diff] are of smaller magnitude than those obtained by using EDR [%].

Taken together, evidence presented in this section suggests that the negative effects of exposure to famine have not been experienced equally across all age cohorts, and that individuals who were older at the time when they were exposed to famine might have been less affected. I explore this conjecture more formally in Section 2.5.3.1.

Heterogenous effects by age: convexity analysis

Results presented in the previous subsection suggest that not all age groups have been affected equally by prior exposure to famine which occurred in their provinces. Coupled with evidence shown in Section 2.5.2.1 and Section 2.5.2.2 that suggests that the relation between age and dietary diversity score is convex in nature, in this section I explore the heterogenous effects of exposure to famine on dietary diversity scores, when the non-linear nature of the link between age and dietary diversity score is taken into account. In particular, I estimate Equation (8) for each of the four waves, using *EDR* [%] and *EDR [diff]* as proxies for prior exposure to famine. Results of this estimation are shown in *Table 1.11 and Table 1.12*, respectively.

The estimated coefficient on *EDR* [%] in *Table 1.11* is negative and statistically significant at 1% level in all four waves, except for 2009. It ranges between -0.119 (2.38 food groups) in 2006 and -0.199 (3.98 food groups) in 2011. The estimated coefficient on *Age* is positive in all four waves, while being statistically significant at 10% level in 2004 and 2006 wave. The estimated coefficient on the quadratic term Age^2 is negative in 2004, 2006 and 2011, and statistically significant at 10% level in 2006 and 2011 wave, and at 1% in 2004 wave. Interestingly, the estimated coefficient on the interaction term $EDR\ x\ Age_h^2$ is mostly positive (with the exception of 2009), and statistically significant at 5% level in 2004 and 2011.

Turning our attention to *Table 1.12*, the estimated coefficient on *EDR[diff]* in *Table 1.12* is negative and statistically significant at 1% level in all four waves, with the exception of 2009 when it's statistically significant at 10% level. . It ranges between -0.005 (0.1 food groups) in 2009 and -0.012 (0.24 food groups) in 2011. The estimated coefficient on *Age* is positive in all four waves, while being statistically significant at 10% level in 2004 and 2006 wave. The estimated coefficient on the quadratic term Age^2 is negative in all four waves, and statistically significant at 10% level in 2006, and at 1% in 2004 wave. Interestingly, the estimated coefficient on the interaction term $EDR\ Age_h^2$ is mostly positive (with the exception of 2009), and statistically significant at 5% level in 2004 and 2011.

Comparing to pooled regressions, we can see that the magnitude of the coefficients are either similar or slightly lower in cross-sectional regressions, yet the statistical significance is consistently higher in the pooled regressions.

2.5.4 Alternative measures of dietary diversity score

In the previous subsection, my analysis focused on estimating the main drivers of dietary diversity score and its relationship with prior exposure to famine, using province-level excess death rates in the period 1959-1961 as proxies for famine exposure. In that analysis, dietary diversity score $D_{h,j}$ was computed using 20 different food groups, as explained in detail in Section *"Methodology and empirical strategy"*. In this subsection, I analyse this relationship be using alternative measures of $D_{h,j}$, using 6 (subsection 2.5.4.1) and 12 (subsection 2.5.4.2) different food groups. The justification for this is grounded in the fact that different literature uses different aggregation level of the food groups, and this dissertation explores whether the results are sensitive to the choice of level of aggregation.

2.5.4.1 Measuring dietary diversity score using 6 food groups

In this subsection, I employ the classification scheme described in *Dietary Pattern* (Sub-section 2.4), and compute dietary diversity score $D_{h,j}$, using 6 food groups. We then re-estimate Equations (4) using this new measure of $D_{h,j}$. In this case, 20 food groups from the previous section have been merged into 6 groups, based on their nutritional characteristics. Therefore, the interpretation of the coefficients is different. While in the previous section 1 food group correspond to coefficient value 0.05, in this case 1 food group corresponds to coefficient value 0.17. In that sense, magnitudes in this section are not comparable to the magnitudes in the previous section. Having said that, statistical significance and the coefficient sign are comparable.

Table 1.6 shows the results of estimating Equation (4), using $D_{h,j}$ computed based on 6 food groups as the dependent variable, where the four waves {2004, 2006, 2009 and 2011} have been pooled together. Column 1 and 3 proxy for exposure to famine using *EDR [%]*, while column 2 and 4 employ *EDR[diff].* The estimated coefficient on *EDR [%]* in Column 1 is negative and statistically significant at 1% level (-0.107), which corresponds to 0.6 food groups. The estimated coefficient on *EDR[diff]* in Column 2 is also negative and statistically significant at 1% level (-0.008), and that translates to 0.05 food groups..

The estimated coefficient on *Age* is positive and significant at 10% level in both Column 3 and Column 4, suggesting that older individuals have higher dietary diversity scores.

When I add the quadratic age term *Age*^{2} into the specification, as shown in Column 3 and Column 4, the estimated coefficients on EDR (EDR[diff], respectively) remain similar, negative and significant at 1% level (-0.104, and -0.008 respectively). The estimated coefficient on the quadratic age term *Age*^{2} is negative and significant at 10% level in both columns (-0.422 and -0.425 respectively), indicating the presence of a convex relationship between age and dietary diversity score. These results suggests that the relation between age and dietary diversity score is not linear, and that much older individuals actually have lower dietary diversity scores, suggesting the presence of a hump-shaped relation between age and dietary diversity scores.

Table 1. 6: Effects of famine (EDR [%/diff]) on Dietary Diversity Score; Pooled regressions (2004-2011)

*Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using **6 food groups***
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include wave dummies.*

	(1)	(2)	(3)	(4)
	Dietary Diversity Score			
log (HH total income)	0.069***	0.068***	0.069***	0.068***
	[12.716]	[12.643]	[12.615]	[12.537]
Education	0.004***	0.004***	0.004***	0.004***
	[4.812]	[4.584]	[4.762]	[4.530]
HH size	-0.010***	-0.009***	-0.011***	-0.010***
	[-4.687]	[-4.251]	[-4.952]	[-4.535]
EDR [%]	-0.107***		-0.104***	
	[-9.619]		[-9.342]	
Age	0.023	0.027	0.514*	0.522*
	[0.664]	[0.795]	[1.775]	[1.816]
Gender	-0.017	-0.018	-0.016	-0.017
	[-1.465]	[-1.527]	[-1.410]	[-1.476]
Han nationality	-0.009	-0.010	-0.007	-0.008
	[-1.093]	[-1.271]	[-0.869]	[-1.069]
Old home food habits	0.002	0.006	0.002	0.006
	[0.146]	[0.399]	[0.142]	[0.394]
Marital Status	0.060***	0.058***	0.059***	0.057***
	[4.253]	[4.134]	[4.145]	[4.025]
Urban	0.017*	0.019*	0.017*	0.018*
	[1.700]	[1.866]	[1.688]	[1.854]
EDR [diff]		-0.008***		-0.008***
		[-11.060]		[-10.832]
Age^2			-0.422*	-0.425*
			[-1.703]	[-1.727]
Wave Dummies	Y	Y	Y	Y
Observations	3,283	3,283	3,283	3,283
R-squared	0.173	0.180	0.174	0.181
AIC	-3,062	-3,090	-3,063	-3,091
BIC	-2,977	-3,004	-2,972	-3,000

2.5.4.1.1 Dietary diversity score using 6 food groups: heterogenous effects

Results presented in the previous subsection suggest that not all age groups have been affected equally by prior exposure to famine, and that this assertion still holds when using an alternative measure of dietary diversity score based on 6 food groups. Coupled with evidence shown in Section 2.5.2.1 and Section 2.5.2.2 that suggests that the relation between age and dietary diversity score is convex in nature, in this section I explore the heterogenous effects of exposure to famine on dietary diversity scores, when the non-linear nature of the link between age and dietary diversity score is taken into account. In particular, I estimate Equation (9) using a pooled dataset consisting of four waves {2003, 2006, 2009 and 2011}, using *EDR* and *EDR[diff]* as proxies for prior exposure to famine. Results of this estimation are shown in *Table 1.7*.

Table 1. 7: Effects of interaction famine of (EDR [%/diff]) and Age/Age^2 on Dietary Diversity Score; Pooled regressions (2004-2011)

Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using ***6 food groups***
Robust standard errors are adjusted for heteroscedasticity; *significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively*; *t-statistics in brackets*. All regressions include wave dummies, while regressions (5)-(8) include both wave and province dummies.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dietary Diversity Score							
log (HH total income)	0.069***	0.068***	0.069***	0.068***	0.061***	0.061***	0.060***	0.060***
	[12.723]	[12.646]	[12.617]	[12.540]	[11.396]	[11.428]	[11.242]	[11.282]
Education	0.004***	0.004***	0.004***	0.004***	0.004***	0.004***	0.004***	0.004***
	[4.812]	[4.578]	[4.761]	[4.530]	[4.218]	[4.247]	[4.163]	[4.198]
HH size	-0.010***	-0.009***	-0.011***	-0.010***	-0.003	-0.003	-0.004*	-0.004*
	[-4.680]	[-4.242]	[-4.945]	[-4.528]	[-1.537]	[-1.549]	[-1.918]	[-1.923]
EDR [%]	-0.102**		-0.109***					
	[-2.002]		[-3.948]					
Age	0.038	0.033	0.522*	0.531*	-0.328**	-0.111*	0.520*	0.507*
	[0.243]	[0.540]	[1.773]	[1.813]	[-2.151]	[-1.828]	[1.833]	[1.784]
EDR [%] x Age	-0.009				0.169**			
	[-0.103]				[2.017]			
Gender	-0.017	-0.018	-0.016	-0.017	-0.015	-0.016	-0.015	-0.015
	[-1.467]	[-1.530]	[-1.398]	[-1.464]	[-1.387]	[-1.409]	[-1.306]	[-1.328]
Han nationality	-0.009	-0.010	-0.007	-0.008	-0.007	-0.007	-0.005	-0.005
	[-1.099]	[-1.281]	[-0.859]	[-1.054]	[-0.797]	[-0.825]	[-0.548]	[-0.576]
Old home food habits	0.002	0.006	0.002	0.006	0.026*	0.026*	0.026*	0.026*
	[0.147]	[0.400]	[0.141]	[0.392]	[1.687]	[1.678]	[1.681]	[1.670]
Marital Status	0.060***	0.058***	0.059***	0.057***	0.052***	0.052***	0.051***	0.051***
	[4.233]	[4.111]	[4.146]	[4.028]	[3.833]	[3.811]	[3.743]	[3.727]
Urban	0.017*	0.018*	0.017*	0.019*	0.043***	0.043***	0.044***	0.043***
	[1.683]	[1.844]	[1.691]	[1.857]	[4.331]	[4.294]	[4.339]	[4.302]
EDR [diff]		-0.007**		-0.008***				
		[-2.251]		[-4.494]				
EDR [diff] x Age		-0.001				0.009		
		[-0.135]				[1.617]		
Age^2			-0.452	-0.440*			-0.764***	-0.543**
			[-1.478]	[-1.674]			[-2.588]	[-2.134]
EDR [%] x Age^2			0.013				0.165**	
			[0.179]				[2.271]	
EDR [diff] x Age^2				0.001				0.009*
				[0.182]				[1.900]
Wave Dummies	Y	Y	Y	Y	Y	Y	Y	Y
Province Dummies					Y	Y	Y	Y
Observations	3,283	3,283	3,283	3,283	3,283	3,283	3,283	3,283
R-squared	0.173	0.180	0.174	0.181	0.250	0.250	0.251	0.251
AIC	-3,060	-3,088	-3,061	-3,089	-3,368	-3,367	-3,370	-3,369
BIC	-2,969	-2,996	-2,964	-2,992	-3,240	-3,239	-3,236	-3,235

Columns 1-4 show the results of estimating Equation (9) with wave dummies, and columns 5-8 show the results of estimating Equation (9) with wave and province dummies. Odd columns show results with *EDR* as a proxy for exposure to famine, while even columns use *EDR [diff].*

As we can see from Column 1, the estimated coefficient on *EDR* is negative and statistically significant at 5% level (-0.102), which translates to 0.6 food groups and confirms my prior results. Interestingly, the estimated coefficient on the interaction term *EDR x Age (Head)* is now negative, yet imprecisely estimated. As Column 2 shows, the estimated coefficient on *EDR] diff]* is negative and statistically significant at 5% level (-0.007), similar to those presented in *Table 1.8*. Interestingly, the estimated coefficient on the interaction term *EDR [diff] x Age (Head)* is now negative, and not precisely estimated.

In Column 3, I add the interaction between exposure to famine and the quadratic age term, *EDR x* Age^2 *(Head).* The estimated coefficient on this interaction term is positive but not statistically significant at conventional levels. In Column 4 I instead use the interaction between exposure to famine and the quadratic age term, *EDR [diff] x* Age^2 *(Head).* The estimated coefficient on this interaction term is also positive and statistically insignificant at conventional levels.

Columns 5-8 show the results of my most saturated pooled regressions, which include wave and province fixed effects. Note that in these specifications, I do not include *EDR* (*EDR [diff],* respectively), since they are time-invariant at the province level, and due to inclusion of province fixed effects they cannot be estimated. As we can see from Column 5, the estimated coefficient on the interaction term *EDR x Age (Head)* is positive, and statistically significant at 5% level (0.169). amd that translates to 1 food group. Similarly, in Column 6, the estimated coefficient on the on the interaction term *EDR [diff] x Age (Head)* is positive, and but not statistically significant (0.009).

In Column 7, I add the interaction between exposure to famine and the quadratic age term, *EDR x* Age^2 *(Head).* The estimated coefficient on this interaction term is positive and statistically significant at 5% level (0.165), and larger in magnitude than the one reported in Column 3, where we did not control for province fixed effects. In Column 8 I instead use the interaction between exposure to famine and the quadratic age term, *EDR [diff] x* Age^2 *(Head).* The estimated coefficient on this interaction term is positive and statistically significant at 10% level (0.009). Once again, both of these estimates are similar in magnitude and sign to those reported in *Table 1.13*.

Taken together, evidence presented in this section suggests that the negative effects of exposure to famine have not been experienced equally across all age cohorts, and that individuals who were older at the time when they were exposed to famine were significantly less effected. As shown in this subsection, this result is robust to alternative measurement of dietary diversity score using 6 food groups.

2.5.4.2 Measuring dietary diversity score using 12 food groups

In this subsection, I employ the classification scheme described in Section *"Methodology and empirical strategy"* and compute dietary diversity score $D_{h,j}$, using 12 food groups. I then re-estimate Equations (1-4) using this new measure of $D_{h,j}$. As described earlier, the magnitude of coefficients in the 12-food groups framework is not comparable to the ones used in 20- and 6-food groups frameworks. In case of 12-food groups, 1 food group corresponds to the coefficient 0.08. Having said that, statistical significance and the coefficient sign are comparable.

Table 1. 8: Effects of famine (EDR [%/diff]) on Dietary Diversity Score; Pooled regressions (2004-2011)

Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using ***12 food groups***
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include wave dummies.*

	(1)	(2)	(3)	(4)
	Dietary Diversity Score			
log (HH total income)	0.066***	0.066***	0.066***	0.065***
	[14.285]	[14.238]	[14.159]	[14.110]
Education	0.002***	0.002***	0.002***	0.002**
	[2.858]	[2.597]	[2.795]	[2.530]
HH size	-0.008***	-0.007***	-0.009***	-0.008***
	[-4.505]	[-4.025]	[-4.939]	[-4.473]
EDR [%]	-0.082***		-0.079***	
	[-8.990]		[-8.643]	
Age	-0.003	0.000	0.598**	0.596**
	[-0.108]	[0.004]	[2.510]	[2.512]
Gender	-0.000	-0.001	0.001	-0.000
	[-0.017]	[-0.098]	[0.065]	[-0.023]
Han nationality	-0.025***	-0.027***	-0.023***	-0.026***
	[-3.894]	[-4.351]	[-3.535]	[-4.034]
Old home food habits	0.019	0.023*	0.019	0.023*
	[1.607]	[1.909]	[1.609]	[1.910]
Marital Status	0.048***	0.046***	0.047***	0.045***
	[4.284]	[4.139]	[4.124]	[3.981]
Urban	0.018**	0.020**	0.018**	0.019**
	[2.001]	[2.163]	[1.986]	[2.150]
EDR [diff]		-0.006***		-0.006***
		[-10.626]		[-10.344]
Age^2			-0.516**	-0.512**
			[-2.535]	[-2.522]
Wave Dummies	Y	Y	Y	Y
Observations	3,283	3,283	3,283	3,283
R-squared	0.159	0.167	0.161	0.169
AIC	-4,259	-4,290	-4,265	-4,297
BIC	-4,167	-4,199	-4,168	-4,200

Table 1.8 shows he results of estimating Equation (5), using $D_{h,j}$ computed based on 12 food groups as the dependent variable, where the four waves {2004, 2006, 2009 and 2011} have been pooled together. Column 1 and 3 proxy for exposure to famine using *EDR*, while column 2 and 4 employ *EDR [diff].* The estimated coefficient on *EDR [%]* in Column 1 is negative and statistically significant at 1% level (-0.082), which translates to 0.99 food groups. The estimated coefficient on *EDR [diff]* in Column 2 is also negative and statistically significant at 1% level (-0.006), and that translates to 0.07 food groups.

The estimated coefficient on *Age* is positive and significant at 10% level in both Column 3 and Column 4, suggesting that older individuals have higher dietary diversity scores.

When I add the quadratic age term *Age^2* into the specification, as shown in Column 3 and Column 4, the estimated coefficients on *EDR* (*EDR [diff],* respectively) remain similar, negative and significant at 1% level (-0.079, and -0.006 respectively). The estimated coefficient on the quadratic age term *Age^2* is negative and significant at 5% level in both columns (-0.516 and -0.512 respectively), indicating the presence of a convex relationship between age and dietary diversity score. These results suggest that the relation between age and dietary diversity score is not linear, and that much older individuals actually have lower dietary diversity scores, suggesting the presence of a hump-shaped relation between age and dietary diversity scores.

2.5.4.2.1 Dietary diversity score using 12 food groups: heterogenous effects

Results presented in the previous subsection suggest that not all age groups have been affected equally by priori exposure to famine, and that this assertion still holds when using an alternative measure of dietary diversity score based on 6 food groups. Coupled with evidence shown in Section 2.5.2.1 and Section 2.5.2.2 that suggests that the relation between age and dietary diversity score is convex in nature, in this section we explore the heterogenous effects of exposure to famine on dietary diversity scores, when the non-linear nature of the link between age and dietary diversity score is taken into account. In particular, I estimate Equation (9) using a pooled dataset consisting of four waves {2003, 2006, 2009 and 2011}, using *EDR* and *EDR [diff]* as proxies for prior exposure to famine. Results of this estimation are shown in *Table 1.9*.

Table 1. 9: Effects of interaction famine of (EDR [%/diff]) and Age/Age² on Dietary Diversity Score; Pooled regressions (2004-2011)

Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using ***12 food groups***
Robust standard errors are adjusted for heteroscedasticity; *significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively*; *t-statistics in brackets*. All regressions include wave dummies, while regressions (5)-(8) include both wave and province dummies.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Dietary Diversity Score				
log (HH total income)	0.066***	0.066***	0.065***	0.065***	0.055***	0.055***	0.054***	0.054***
	[14.260]	[14.231]	[14.106]	[14.076]	[12.278]	[12.321]	[12.060]	[12.114]
Education	0.002***	0.002***	0.002***	0.002**	0.002***	0.002***	0.002***	0.002***
	[2.856]	[2.625]	[2.778]	[2.550]	[2.747]	[2.794]	[2.666]	[2.723]
HH size	-0.008***	-0.007***	-0.009***	-0.008***	-0.001	-0.001	-0.002	-0.002
	[-4.536]	[-4.072]	[-5.033]	[-4.605]	[-0.397]	[-0.421]	[-1.060]	[-1.086]
EDR [%]	-0.143***		-0.123***					
	[-3.289]		[-5.185]					
Age	-0.184	-0.075	0.673***	0.694***	-0.382***	-0.143***	0.689***	0.686***
	[-1.365]	[-1.430]	[2.790]	[2.890]	[-2.869]	[-2.757]	[2.938]	[2.924]
EDR [%] x Age	0.103				0.196***			
	[1.399]				[2.664]			
Gender	0.000	-0.000	0.002	0.001	-0.007	-0.007	-0.006	-0.006
	[0.050]	[-0.017]	[0.161]	[0.092]	[-0.794]	[-0.810]	[-0.665]	[-0.679]
Han nationality	-0.025***	-0.026***	-0.022***	-0.024***	-0.000	-0.000	0.003	0.002
	[-3.819]	[-4.203]	[-3.420]	[-3.824]	[-0.029]	[-0.061]	[0.373]	[0.340]
Old home food habits	0.019	0.023*	0.019	0.022*	0.033***	0.033***	0.033***	0.033***
	[1.594]	[1.892]	[1.597]	[1.895]	[2.874]	[2.857]	[2.884]	[2.863]
Marital Status	0.049***	0.048***	0.048***	0.047***	0.044***	0.044***	0.043***	0.043***
	[4.372]	[4.268]	[4.235]	[4.135]	[4.058]	[4.049]	[3.922]	[3.919]
Urban	0.019**	0.021**	0.019**	0.021**	0.046***	0.046***	0.047***	0.046***
	[2.102]	[2.299]	[2.136]	[2.342]	[5.110]	[5.090]	[5.135]	[5.118]
EDR [diff]		-0.011***		-0.009***				
		[-3.998]		[-6.206]				
EDR [diff] x Age		0.009*				0.012**		
		[1.795]				[2.448]		
Age^2			-0.798***	-0.683***			-0.975***	-0.732***
			[-3.136]	[-3.162]			[-3.911]	[-3.461]
EDR [%] x Age^2			0.124*				0.199***	
			[1.917]				[3.105]	
EDR [diff] x Age^2				0.010**				0.012***
				[2.355]				[2.927]
Wave Dummies	Y	Y	Y	Y	Y	Y	Y	Y
Province Dummies					Y	Y	Y	Y
Observations	3,283	3,283	3,283	3,283	3,283	3,283	3,283	3,283
R-squared	0.160	0.168	0.162	0.170	0.240	0.239	0.242	0.242
AIC	-4,259	-4,290	-4,265	-4,297	-4,575	-4,574	-4,583	-4,582
BIC	-4,167	-4,199	-4,168	-4,200	-4,446	-4,446	-4,449	-4,448

Columns 1-4 show the results of estimating Equation (9) with wave fixed effects, and columns 5-8 show the results of estimating Equation (9) with wave and province fixed effects. Odd columns show results with *EDR* as a proxy for exposure to famine, while even columns use *EDR [diff].*

As we can see from Column 1, the estimated coefficient on *EDR* is negative and statistically significant at 1% level (-0.143), which translates to 1.7 food groups, confirming our prior results. Interestingly, the estimated coefficient on the interaction term *EDR x Age (Head)* is positive, yet imprecisely estimated. As Column 2 shows, the estimated coefficient on *EDR [diff]* is negative and statistically significant at 1% level (-0.011), and that corresponds to 0.13 food groups. Interestingly, the estimated coefficient on the interaction term *EDR [diff] x Age (Head)* is positive, and statistically significant at 10% level.

In Column 3, I add the interaction between exposure to famine and the quadratic age term, *EDR x* Age_h^2. The estimated coefficient on this interaction term is positive and statistically significant at 10% level (0.124). In Column 4 I instead use the interaction between exposure to famine and the quadratic age term, *EDR [diff] x* Age_h^2. The estimated coefficient on this interaction term is also positive and statistically significant at 10% level (0.010).

Columns 5-8 show the results of my most saturated pooled regressions, which include wave and province fixed effects. Note that in these specifications, I do not include *EDR* (*EDR [diff],* respectively), since they are time-invariant at the province level, and due to inclusion of province fixed effects they cannot be estimated. As we can see from Column 5, the estimated coefficient on the interaction term *EDR x Age (Head)* is positive, and statistically significant at 1% level (0.196). Similarly, in Column 6, the estimated coefficient on the on the interaction term *EDR [diff] x Age (Head)* is positive, and statistically significant at 5% level (0.012).

In Column 7, we add the interaction between exposure to famine and the quadratic age term, *EDR x* Age^2 *(Head).* The estimated coefficient on this interaction term is positive and statistically significant at 1% level (0.199), and larger in magnitude than the one reported in Column 3, where I did not control for province fixed effects. In Column 8 I instead use the interaction between exposure to famine and the quadratic age term, *EDR [diff] x* Age_h^2. The estimated coefficient on this interaction term is positive and statistically significant at 1% level (0.012.

Taken together, evidence presented in this section suggests that the negative effects of exposure to famine have not been experienced equally across all age cohorts, and that individuals who were older at the time when they were exposed to famine were significantly less effected. As shown in this subsection, this result is robust to alternative measurement of dietary diversity score using 12 and 6 food groups.

2.5.5 Birth cohort analysis

In this subsection, I explore whether any particular birth-year cohort around period of famine is being more affected than others, measured by DDS. I first quantify the lasting effects of the famine on dietary diversity score of the survivors by estimating Equation (10):

Note that k=1962 refers to a birth cohort in gestation in 1961 and born in 1962. I include this birth cohort because, as discussed in the previous section, the exposure to famine during the foetal period may exert significant effects on body composition and growth after birth, driven either by epigenetics or by food consumption patterns or both.

The coefficient of the interaction between the excess death rate and birth cohort dummy variables measures the causal effect of the famine on dietary diversity score. I expect that the magnitude

of these estimated coefficients varies with birth cohorts. More specifically, based on the earlier analysis, I expect a larger impact of the famine on the relatively young birth cohorts, especially those who were in early childhood during the famine. The total sample includes those individuals born between 1954 and 1962 as the treatment group, and individuals born between 1963 and 1967, which are treated as the control group.

Columns (1)–(4) of *Table 1.10* present the cross-sectional regression results estimating the long-term effect of famine on dietary diversity score later in life, whereby each column presents the results of cross-sectional regressions for each four waves {2004, 2006, 2009 and 2011}. Column 5 represents results from pooled regression, which includes the four waves. Following (Y. Chen & Zhou, 2007) the dummy variables for birth cohorts from 1954 to 1962 are controlled in the regressions in Table 16, but their coefficients are not reported due to space limitations. I mainly report the coefficients of interaction terms between excess death rate *EDR* [%] and birth cohorts, which, as discussed above, measure the estimated effects of the famine on dietary diversity score for each birth cohort. The estimated coefficient on EDR is negative in all four columns, but statistically significant at 1% level only for the 2004 wave (-0.074).

Table 1. 10: Effects of Famine (EDR [%]) on Dietary Diversity Score of birth cohorts 1954-1962; Cross sectional and Pooled regressions (2004-2011)

Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); Dependent variable Dietary Diversity Score is calculated using ***20 food groups*** *Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include birth-cohort dummy (suppressed for brevity).*

	Dietary Diversity Score				
	2004	2006	2009	2011	2004-2011
log (HH total income)	0.055***	0.083***	0.049***	0.050***	0.071***
	[3.258]	[5.358]	[3.271]	[4.181]	[11.652]
Education	0.001	0.004*	0.003	-0.001	0.001
	[0.599]	[1.799]	[1.039]	[-0.302]	[1.069]
HH size	-0.013**	-0.022***	-0.004	-0.011*	-0.014***
	[-2.171]	[-3.952]	[-0.668]	[-1.701]	[-4.606]
Gender	-0.029	0.045	0.110***	0.009	0.025
	[-0.889]	[1.385]	[4.499]	[0.204]	[1.409]
Han nationality	-0.020	0.009	-0.031	-0.012	-0.014
	[-1.071]	[0.452]	[-1.355]	[-0.665]	[-1.374]
Old home food habits	-0.037	0.016	-0.083*	0.054	-0.020
	[-0.983]	[0.481]	[-1.859]	[1.349]	[-0.993]
Marital Status	0.085**	0.019	0.107*	0.008	0.046**
	[2.585]	[0.424]	[1.861]	[0.170]	[2.251]
Urban	0.019	0.009	0.037	0.099***	0.032**
	[0.992]	[0.298]	[0.964]	[3.077]	[2.284]
EDR [%]	-0.074*	-0.073	-0.019	-0.053	-0.057*
	[-1.678]	[-1.177]	[-0.350]	[-0.619]	[-1.812]
BirthYear1954 x EDR [%]	-0.020	-0.026	-0.070	0.052	-0.016
	[-0.289]	[-0.361]	[-1.018]	[0.424]	[-0.398]
BirthYear1955 x EDR [%]	0.015	-0.038	-0.085	-0.099	-0.049
	[0.225]	[-0.450]	[-1.166]	[-1.056]	[-1.218]
BirthYear1956 x EDR [%]	-0.035	-0.049	-0.053	-0.030	-0.042
	[-0.689]	[-0.656]	[-0.564]	[-0.321]	[-1.064]
BirthYear1957 x EDR [%]	0.035	0.076	-0.283**	-0.152*	-0.060
	[0.503]	[1.056]	[-2.086]	[-1.677]	[-1.325]
BirthYear1958 x EDR [%]	0.002	0.053	-0.050	0.040	0.026
	[0.033]	[0.367]	[-0.201]	[0.317]	[0.443]
BirthYear1959 x EDR [%]	-0.071	-0.042	0.129	-0.031	-0.051
	[-0.728]	[-0.503]	[0.418]	[-0.182]	[-0.964]
BirthYear1960 x EDR [%]	0.005	0.134*	-0.120	0.126	0.056
	[0.065]	[1.750]	[-1.283]	[1.019]	[0.981]
BirthYear1961 x EDR [%]	-0.008	0.032	-0.078	0.025	-0.010
	[-0.102]	[0.198]	[-0.845]	[0.265]	[-0.194]
BirthYear1962 x EDR [%]	0.036	-0.071	0.302	-0.197	0.001
	[0.291]	[-0.578]	[1.271]	[-1.376]	[0.012]
Birth Cohort Dummy	Y	Y	Y	Y	Y
Observations	327	264	239	234	1,064
R-squared	0.160	0.278	0.219	0.272	0.212
AIC	-768.6	-585.9	-462.8	-478.6	-1,754
BIC	-646.6	-469.7	-349.8	-367.3	-1,615

Most famine-affected birth cohorts have lower dietary diversity scores than the counterfactual case that the famine had not occurred, which is reflected in the negative coefficients of the interaction terms *EDRxBirthYear*. This implies that the famine generally caused adverse effect on the dietary diversity scores of the survivors. Among the cohorts with adverse effects of famine, the birth cohorts of 1957 have more severe and also statistically significant effects. As evidenced in Column 3 and Column 4, the estimated coefficient on the interaction term is negative and statistically significant at 5% (1%) level respectively. As per wave 2009, cohort born in 1975 consume 5.6 food groups less than those born after the famine, while wave 2011 suggests that this difference is 3 food groups. These results suggest that exposure to famine during early years of childhood (i.e. for children who were around 2 years of age when the famine started) gives rise to more devastating long-term effects in later life, which is consistent with general findings in population health literature (e.g., Barker, 1989, 1992; Heymann et al., 2005).

In *Table 1.12* I repeat this analysis, but now I use *EDR [diff]* as the measure of famine exposure. Similar to *Table 1.18*, columns 1-4 show results for the four cross-sectional regressions for each of the waves {2004, 2006, 2009 and 2011}. Again, I find similar results as shown in *Table 1.18*. In wave 2009, thise bornin 1957 on average consumed 0.4 food groups less than those born after the Great Famine.

Both Table 1.10 and 1.11 also show the results of the pooled regression (in column 5). While the estimated coefficient on *EDR* is negative and statistically significant at 10% (5%) level respectively, most *EDR [%] x BirthYear* coefficients are negative, although imprecisely estimated.

Taken together, these results demonstrate large and devastating effects of the great famine, which occurred decades ago, on the surviving population. They also lend strong support to the notion that exposure to adverse health shocks in early childhood as well as in the prenatal period translates into serious health consequences in adulthood.

My estimated results show that the 1957 birth cohort suffered most severely, followed by the individuals born in 1956 and 1955, although the differences between them seem to be very small. These year cohorts were between two and five years old when the famine occurred. It is also important to bear in mind that these results are obtained on relatively small sample sizes, (around 250 observations for cross-sectional regressions and 1064 for the pooled regression). Hence, the power of this analysis can be a potential concern.

Table 1. 11: Effects of Famine (EDR [diff]) on Dietary Diversity Score of birth cohorts 1954-1962; Cross sectional and Pooled regressions (2004-2011)

*Variables are defined as described in Table 1.2. Excess Death Rate EDR [diff] calculated as the difference between 3 famine years and 3 pre-famine years.; Dependent variable Dietary Diversity Score is calculated using **20 food groups** Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include birth-cohort dummy (suppressed for brevity).*

	Dietary Diversity Score				
	2004	2006	2009	2011	2004-2011
log (HH total income)	0.056***	0.083***	0.047***	0.052***	0.071***
	[3.351]	[5.543]	[3.142]	[4.467]	[11.739]
Education	0.001	0.004*	0.003	-0.000	0.001
	[0.425]	[1.776]	[0.936]	[-0.115]	[0.975]
HH size	-0.012**	-0.020***	-0.003	-0.011*	-0.013***
	[-1.993]	[-3.622]	[-0.484]	[-1.743]	[-4.272]
Gender	-0.026	0.044	0.110***	0.005	0.025
	[-0.813]	[1.383]	[4.271]	[0.128]	[1.391]
Han nationality	-0.020	0.010	-0.033	-0.003	-0.012
	[-1.104]	[0.463]	[-1.537]	[-0.190]	[-1.277]
Old home food habits	-0.031	0.023	-0.079*	0.054	-0.017
	[-0.842]	[0.704]	[-1.898]	[1.373]	[-0.831]
Marital Status	0.082**	0.011	0.102*	0.008	0.043**
	[2.507]	[0.261]	[1.934]	[0.169]	[2.104]
Urban	0.022	0.007	0.043	0.095***	0.033**
	[1.136]	[0.252]	[1.122]	[2.906]	[2.379]
EDR[Diff]	-0.006**	-0.005	-0.003	-0.003	-0.005**
	[-2.175]	[-1.111]	[-0.781]	[-0.636]	[-2.294]
BirthYear1954 x EDR [diff]	-0.001	-0.004	-0.006	0.002	-0.002
	[-0.231]	[-0.769]	[-1.464]	[0.303]	[-0.819]
BirthYear1955 x EDR [diff]	0.002	-0.002	-0.004	-0.006	-0.002
	[0.352]	[-0.387]	[-0.907]	[-0.990]	[-0.930]
BirthYear1956 x EDR [diff]	-0.000	-0.004	-0.003	-0.002	-0.002
	[-0.085]	[-0.827]	[-0.486]	[-0.278]	[-0.744]
BirthYear1957 x EDR [diff]	0.003	0.004	-0.019**	-0.008	-0.003
	[0.633]	[0.787]	[-2.322]	[-1.479]	[-1.207]
BirthYear1958 x EDR [diff]	0.001	0.001	0.000	0.002	0.002
	[0.219]	[0.099]	[0.013]	[0.351]	[0.536]
BirthYear1959 x EDR [diff]	-0.002	-0.002	0.001	-0.002	-0.002
	[-0.406]	[-0.324]	[0.036]	[-0.151]	[-0.758]
BirthYear1960 x EDR [diff]	0.001	0.005	-0.006	0.012*	0.005
	[0.178]	[0.961]	[-1.036]	[1.674]	[1.232]
BirthYear1961 x EDR [diff]	-0.000	0.003	-0.005	0.002	-0.000
	[-0.016]	[0.281]	[-0.814]	[0.327]	[-0.093]
BirthYear1962 x EDR [diff]	0.002	-0.003	0.010	-0.006	0.001
	[0.238]	[-0.286]	[0.569]	[-0.582]	[0.139]
Birth Cohort Dummy	Y	Y	Y	Y	Y
Observations	327	264	239	234	1,064
R-squared	0.169	0.283	0.225	0.264	0.218
AIC	-563.7	-422.8	-326.7	-383.7	-1,763
BIC	-457.6	-322.7	-229.4	-286.9	-1,624

2.5.6 Early life famine and dietary diversity score: heterogenous effect by income

2.5.6.1 Cross-sectional analysis

I begin my analysis by estimating Equation (12) using the CHNS dataset covering 4 survey waves (2004, 2006, 2009 and 2011). *Table 1.20* shows the results of cross-sectional OLS regressions for each wave, where the dependent variable $D_{h,j}$ denotes average dietary diversity score of household *h* located in province *j*. $D_{h,j}$ is computed using the 20 food groups from the Food Composition Tables described in *Dietary pattern* part of 2.4 section.

Columns (1), (3), (5) and (7) show the results of estimating Equation (12), while Columns (2), (4), (6) and (8) show the results of estimating Equation (12.1) for each wave {2004, 2006, 2009 and 2011}. The vector of independent variables includes: natural logarithm of household total income, household size, (household head's) completed years of education, age, gender and marital status of the head of the household, (household head's) nationality, whether household is located in an urban or rural environment, and whether they maintain food habits from their previous home. Specifications reported in even columns include province dummy/fixed effects, while all specifications include robust standard errors that are adjusted for heteroscedasticity.

The independent variable of interest *EDR_inv* captures the inverse of the normalized excess death rate *EDR [%]* between time period 1959-1961 in province *j*, where the household is located. These variables as such serve as a proxy for the household's previous exposure to famine Xu et al. (2016) . As we can see from *Table 1.20*, the estimated coefficients on the inverse excess death rate *EDR_inv* are positive and significant at 5% level in all 2006 and 2009, and significant at 10% level in all 2004. These results confirm my prior results from *Section Two* and suggest that households that were less exposed to Great China Famine, as proxied by the higher inverse excess death rates between 1959-1961, tend to have higher average dietary diversity scores later in life.

The main coefficient of interest on the interaction term *EDR_inv x log(HH total income)* is positive and imprecisely estimated in all waves, with the exception of 2004 (Column 1), where it is significant at 5% level. This result confirms my intuition that households with higher disposable income who were exposed to famine early on in their lives, or whose parents have been exposed tend to have higher dietary diversity score later on. In Annex 1, *Table A.1.12* I repeat this analysis by using *EDR_inv[diff]* which captures the inverse of the normalized excess death rate *EDR[diff]* between time period 1959-1961 and 1956-1958 in province *j*, where the household is located. Results remain quantitatively and qualitatively similar. It is worth mentioning here that while most of the estimated coefficients on the *EDR_inv x log(HH total income)* interaction term in *Table 1.12* are not statistically significant possibly due to limited sample size used in the cross-sectional analysis, they are of the anticipated sign. In the next subsection, I conduct the pooled sample analysis with the aim of estimating this relationship by taking into account the wave-specific unobservables that can be driving my results, as well as to increase the regression power.

Table 1. 12: Effects of interaction of Famine inverted term (EDR_inv [%]) and Income on Dietary Diversity Score; Cross sectional regressions

Variables are defined as described in Table 1.2. EDR_inv [%] represents inverted value of Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); Dependent variable Dietary Diversity Score is calculated using 20 food groups.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; Columns (2), (4), (6) and (8) include province dummies.*

	Dietary Diversity Score							
	2004		2006		2009		2011	
log (HH total income)	0.046***	0.057***	0.037**	0.034*	0.025*	0.022*	0.041**	0.040*
	[3.338]	[4.106]	[2.102]	[1.953]	[1.826]	[1.652]	[1.987]	[1.889]
Education	0.001	0.001	0.003**	0.003**	0.002	0.002*	0.001	-0.000
	[1.144]	[1.255]	[2.580]	[2.395]	[1.619]	[1.762]	[0.363]	[-0.066]
EDR_inv [%]	0.038*		0.061**		0.074**		0.049	
	[1.677]		[2.244]		[2.144]		[1.047]	
Age	0.011	-0.012	0.091**	0.056	0.006	-0.074	-0.113**	-0.139**
	[0.323]	[-0.352]	[1.971]	[1.248]	[0.104]	[-1.540]	[-2.003]	[-2.424]
EDR_inv [%] x log (HH total income)	0.058**	0.013	0.051	0.043	0.033	0.015	0.041	0.035
	[2.199]	[0.505]	[1.596]	[1.402]	[1.244]	[0.565]	[1.185]	[0.965]
Gender	-0.002	-0.009	-0.014	-0.022	0.018	0.002	0.003	-0.001
	[-0.225]	[-0.876]	[-0.831]	[-1.345]	[1.077]	[0.111]	[0.192]	[-0.094]
HH size	-0.009***	-0.004*	-0.009***	-0.001	-0.005	0.004	-0.002	0.002
	[-3.744]	[-1.902]	[-2.954]	[-0.255]	[-1.583]	[1.642]	[-0.455]	[0.679]
Han nationality	-0.014	0.004	0.002	0.019**	-0.023**	0.013	-0.038***	-0.033**
	[-1.482]	[0.436]	[0.238]	[1.977]	[-2.198]	[1.286]	[-3.262]	[-2.564]
Old home food habits	0.006	0.010	0.024	0.043**	-0.017	-0.003	0.034	0.046*
	[0.315]	[0.538]	[1.319]	[2.385]	[-0.816]	[-0.154]	[1.281]	[1.730]
Marital Status	0.062***	0.059***	0.038**	0.023	0.012	0.012	0.031	0.030
	[4.920]	[4.304]	[2.164]	[1.272]	[0.520]	[0.600]	[1.421]	[1.394]
Urban	0.031***	0.045***	0.018	0.048***	0.050**	0.076***	0.034*	0.050**
	[2.853]	[4.184]	[1.094]	[2.940]	[2.496]	[3.820]	[1.684]	[2.401]
Province Dummies		Y		Y		Y		Y
Observations	1,001	1,001	832	832	764	764	686	686
R-squared	0.166	0.233	0.178	0.292	0.129	0.294	0.189	0.224

2.5.6.2 Pooled cross-sectional analysis

Cross-sectional results presented in the previous part provide suggestive evidence that it is precisely those high-income households that were less severely exposed to famine early on in their lives that have higher dietary diversity scores. In this subsection, I extend this analysis by pooling the four waves {2004, 2006, 2009 and 2011} together. In addition, I employ the extended pooled sample that also incorporates the 1997 and 2001 wave.[4] The results of

[4] For detailed data construction description see Dietary Patterns sub-section in 2.4.

estimating Equation (13) and (13.1) on the pooled cross-sectional sample are shown in *Table 1.13*.

Columns (1)-(4) show the results for the 1997-2011 pooled sample, while Columns (5)-(8) show the results for the 2004-2011 pooled sample. As before, all specifications include wave dummies, while even specifications also include province dummies. Estimated coefficient on the interaction term *EDR_inv x log(HH total income)* is positive and significant at 1% level in both Column (1) and (2), 0.069 (1.38 food groups) and 0.056 (1.12 food groups) respectively, suggesting that controlling for wave and province unobserved heterogeneity, higher income households with exposed to famine are those with higher dietary diversity scores. Similarly, estimated coefficient on the interaction term *EDR_inv[diff] x log(HH total income)* is positive and significant at 1% level in both Column (3) and (4), 0.063 (1.26 food groups) and 0.051 (1 food group) respectively, suggesting that higher income households with less prior exposure to famine are precisely those that have higher dietary diversity scores, irrespective of the chosen proxy for prior famine exposure. Results shown in Columns (5)-(8) further corroborate these findings.

2.5.6.3 Pooled split sample analysis

Here, I split the sample into those who were born before and during the Great famine (1954 - 1962) and those born after the famine (1963 – 1988). Results from previous estimations suggest that exposure to famine affects both famine survivors, and those who were born after the famine, but in the areas where famine occurred. In this sub-section, I test whether the effects of famine diminish over time. That is whether the significance level, and potentially sign differ between famine survivors and generations born after the famine. Columns (1)-(4) show the results for the famine cohort (1954-1962) of the pooled sample, while Columns (5)-(8) show the results for the post-famine cohort (1963-1988) of the pooled sample. Odd columns include wave dummies only, while even columns include province dummies too.

Table 1. 13: Effects of interaction of Famine inverted term (EDR_inv [%/diff]) and Income on Dietary Diversity Score; Pooled regression (1997-2011) and (2004-2011)

Variables are defined as described in Table 1.2. EDR_inv [%] represents inverted value of Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR_inv [diff] represents inverted value of Excess Death Rate calculated as difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using 20 food groups.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets. All regressions include wave dummies, while regressions (2), (4), (6) and (8) include both wave and province dummies.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1997-2011				2004-2011			
log (HH total income)	0.024***	0.021***	0.023***	0.022***	0.044***	0.043***	0.044***	0.042***
	[3.477]	[3.061]	[3.357]	[3.096]	[5.487]	[5.280]	[5.375]	[5.104]
Education	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	[4.870]	[4.664]	[4.614]	[4.684]	[2.842]	[2.830]	[2.624]	[2.824]
EDR_inv [%]	-0.009				0.069***			
	[-0.833]				[4.772]			
Age (Head)	0.024	0.001	0.026*	0.000	0.007	-0.029	0.010	-0.029
	[1.595]	[0.044]	[1.690]	[0.030]	[0.287]	[-1.276]	[0.425]	[-1.288]
EDR_inv [%] x log (HH total income)	0.069***	0.056***			0.030**	0.015		
	[5.881]	[4.840]			[2.192]	[1.056]		
Gender	-0.006	-0.009*	-0.006	-0.009*	-0.000	-0.007	-0.000	-0.007
	[-1.210]	[-1.755]	[-1.259]	[-1.750]	[-0.004]	[-1.033]	[-0.046]	[-1.035]
HH size	-0.007***	-0.002**	-0.006***	-0.002**	-0.006***	-0.000	-0.006***	-0.000
	[-6.796]	[-2.412]	[-6.347]	[-2.431]	[-4.444]	[-0.219]	[-4.001]	[-0.226]
Han nationality	-0.008**	0.006	-0.011***	0.006	-0.017***	0.003	-0.017***	0.003
	[-2.031]	[1.565]	[-2.739]	[1.554]	[-3.364]	[0.492]	[-3.441]	[0.485]
Old home food habits	0.003	0.014*	0.005	0.014*	0.009	0.019*	0.011	0.019*
	[0.448]	[1.918]	[0.722]	[1.919]	[0.812]	[1.929]	[1.107]	[1.932]
Marital Status	0.035***	0.032***	0.035***	0.032***	0.038***	0.033***	0.036***	0.033***
	[5.725]	[5.355]	[5.620]	[5.391]	[4.093]	[3.700]	[3.971]	[3.719]
Urban	0.027***	0.041***	0.028***	0.041***	0.032***	0.053***	0.033***	0.053***
	[5.559]	[8.148]	[5.682]	[8.142]	[4.020]	[6.728]	[4.174]	[6.724]
EDR_inv [diff]			0.003				0.073***	
			[0.293]				[5.460]	
EDR_inv [diff] x log (HH total income)			0.063***	0.051***			0.028**	0.015
			[5.767]	[4.689]			[2.153]	[1.176]
Wave Dummies	Y	Y	Y	Y	Y	Y	Y	Y
Province Dummies		Y		Y		Y		Y
Observations	5,748	5,748	5,748	5,748	3,283	3,283	3,283	3,283
R-squared	0.186	0.250	0.190	0.250	0.194	0.271	0.200	0.271
AIC	-10,229	-10,705	-10,258	-10,706	-5,480	-5,797	-5,507	-5,798
BIC	-10,116	-10,552	-10,145	-10,553	-5,389	-5,669	-5,415	-5,670

Variable of interest is the interaction term between EDR and income. As per *Table 1.14*, in case of *EDR [%] x Income*, this coefficient is positive and significant at 5% for those born between 1954 and 1962 (0.052 in Column 1, which translates to 1 food group) and it is positive yet statistically insignificant for those born after the famine, that is between 1963 and 1988 (0.046 in Column 5). Similar result are obtained for the estimated coefficient on the interaction term *EDR [diff] x Income*. As we can see from Column 3., the estimate coefficient for those born between 1954 and 1962 is 0.050 (1 food group) and it is statistically significant at 5% level. For those respondents born after the famine, that is between 1963 and 1988, the estimated coefficient on the interaction term is not statistically different from zero (Column 7).

When province dummy variables are included in regressions, all estimated coefficients on the interaction term *EDR [%] x Income* and *EDR [diff] x Income*, respectively, are imprecisely estimated (Columns 2, 4, 6 and 8), which suggests that unobservable time-invariant province factors can be driving the obtained results.

Taken together, these results suggest that the heterogenous effect of famine on DDS by income level diminishes over generations, and that it is reduced for those born after the famine.

Table 1. 14: Effects of interaction of Famine inverted term (EDR_inv [%/diff]) and Income on Dietary Diversity Score; Pooled regression (2004-2011), split sample Cohort 1954-1962 and Cohort 1963-1988

Variables are defined as described in Table 1.2. EDR_inv [%] represents inverted value of Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR_inv [diff] represents inverted value of Excess Death Rate calculated as difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using 20 food groups.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets. All regressions include wave dummies, while regressions (2), (4), (6) and (8) include both wave and province dummies.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES		Born 1954-1962				Born 1963-1988		
log (HH total income)	0.025	0.033**	0.023	0.029*	0.006	0.006	0.009	0.011
	[1.586]	[2.122]	[1.444]	[1.894]	[0.284]	[0.335]	[0.419]	[0.568]
Education	0.002	0.002*	0.001	0.002*	0.002	0.004	0.003	0.004
	[1.397]	[1.885]	[1.261]	[1.878]	[0.741]	[1.594]	[1.113]	[1.617]
EDR_inv [%]	0.060**				0.063			
	[2.136]				[1.465]			
Age (Head)	0.160	0.183	0.170	0.183	-0.966**	-0.821**	-1.029***	-0.831**
	[1.079]	[1.322]	[1.152]	[1.325]	[-2.516]	[-2.242]	[-2.702]	[-2.255]
EDR_inv [%] x log (HH total income)	0.052**	0.023			0.046	0.047		
	[1.984]	[0.858]			[1.281]	[1.470]		
Gender	0.055***	0.035***	0.055***	0.035***	-0.005	-0.013	-0.004	-0.013
	[3.616]	[2.731]	[3.651]	[2.738]	[-0.229]	[-0.474]	[-0.151]	[-0.468]
HH size	-0.014***	-0.004	-0.013***	-0.004	-0.008**	-0.005	-0.008**	-0.005
	[-4.429]	[-1.082]	[-4.085]	[-1.088]	[-2.128]	[-1.200]	[-2.071]	[-1.235]
Han nationality	-0.012	0.015	-0.009	0.015	-0.020	-0.001	-0.020	-0.000
	[-1.241]	[1.344]	[-0.886]	[1.320]	[-1.181]	[-0.041]	[-1.201]	[-0.025]
Old home food habits	-0.039*	-0.020	-0.034	-0.020	-0.088	-0.065	-0.088	-0.066
	[-1.802]	[-1.017]	[-1.607]	[-1.008]	[-1.447]	[-1.281]	[-1.480]	[-1.294]
Marital Status	0.045**	0.005	0.042**	0.005	0.016	0.005	0.020	0.005
	[2.117]	[0.241]	[2.009]	[0.258]	[0.777]	[0.227]	[0.975]	[0.263]
Urban	0.035**	0.065***	0.036**	0.065***	0.008	0.050*	0.009	0.050*
	[2.374]	[4.290]	[2.484]	[4.282]	[0.352]	[1.905]	[0.430]	[1.922]
EDR_inv [diff]			0.063**				0.074*	
			[2.426]				[1.912]	
EDR_inv [diff] x log (HH total income)			0.050**	0.026			0.038	0.036
			[2.030]	[1.101]			[1.161]	[1.200]
Wave dummies	Y	Y	Y	Y	Y	Y	Y	Y
Province dummies		Y		Y		Y		Y
Observations	918	918	918	918	360	360	360	360
R-squared	0.212	0.303	0.219	0.303	0.205	0.323	0.218	0.322
AIC	-1,531	-1,632	-1,539	-1,632	-631	-677	-637	-676
BIC	-1,458	-1,530	-1,466	-1,531	-572	-595	-578	-595

2.5.7 Early life famine and dietary diversity score: heterogenous effect by education

2.5.7.1 Cross-sectional analysis

In this subsection I test the conjecture that the positive relation between an individual's education level and dietary diversity score is driven by those individuals who had a higher education, when exposed to famine, or when their parents were exposed to. In particular, I estimate Equation (14) on each of the four cross-sections {2004, 2006, 2009 and 2011}. Results are shown in *Table 1.15.*

As before, the estimated coefficient on the inverse normalized excess death rate measure *EDR_inv* is positive and significant at 1% level in all four waves, suggesting that households who had lower prior exposure to famine have higher dietary diversity scores later in life. The estimated coefficient on the *EDR_inv x Compl. Years Education (Head)* interaction term is positive in all four ways, ranging from 0.001 to 0.004, however, it is imprecisely estimated. This result provides weak evidence that among the highly educated households, higher dietary diversity score is driven by those who had lower exposure to famine in their childhood. I further test this conjecture in the pooled regression setting in the next subsection. In Annex 1, *Table A.1.13* I repeat this cross-sectional analysis by using *EDR_inv[diff]* which captures the inverse of the normalized excess death rate *EDR[diff]* between time period 1959-1961 and 1956-1958 in province *j*, where the household is located. Results remain quantitatively and qualitatively similar.

Table 1. 15: Effects of interaction of Famine inverted term (EDR_inv [%]) and Education on Dietary Diversity Score; Cross sectional regressions.

Variables are defined as described in Table 1.2. EDR_inv [%] represents inverted value of Excess Death Rate, calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); Dependent variable Dietary Diversity Score is calculated using 20 food groups.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; regressions (2), (4), (6) and (8) include province dummies.*

	Dietary Diversity Score							
	2004		2006		2009		2011	
log (HH total income)	0.075***	0.064***	0.065***	0.058***	0.043***	0.030***	0.063***	0.058***
	[9.463]	[8.172]	[7.284]	[7.200]	[5.824]	[4.475]	[8.561]	[7.322]
Education	-0.000	-0.000	0.003	0.002	0.001	-0.000	-0.001	-0.001
	[-0.099]	[-0.113]	[1.428]	[1.051]	[0.410]	[-0.073]	[-0.258]	[-0.358]
EDR_inv [%]	0.061**		0.097***		0.099***		0.093***	
	[2.567]		[4.131]		[3.285]		[2.754]	
Age	0.010	-0.013	0.096**	0.059	0.009	-0.072	-0.114**	-0.142**
	[0.278]	[-0.388]	[2.073]	[1.312]	[0.174]	[-1.505]	[-2.011]	[-2.453]
EDR_inv[%] x Education	0.003	0.003	0.001	0.001	0.002	0.004	0.002	0.002
	[0.876]	[0.935]	[0.234]	[0.432]	[0.586]	[1.096]	[0.503]	[0.379]
Gender	-0.002	-0.009	-0.013	-0.021	0.021	0.004	0.003	-0.001
	[-0.156]	[-0.817]	[-0.819]	[-1.330]	[1.254]	[0.271]	[0.222]	[-0.066]
HH size	-0.009***	-0.004*	-0.009***	-0.001	-0.005	0.004	-0.002	0.002
	[-3.727]	[-1.883]	[-2.925]	[-0.232]	[-1.597]	[1.636]	[-0.449]	[0.701]
Han nationality	-0.015	0.004	0.002	0.020**	-0.022**	0.014	-0.038***	-0.032**
	[-1.570]	[0.362]	[0.250]	[2.018]	[-2.096]	[1.361]	[-3.248]	[-2.506]
Old home food habits	0.007	0.010	0.025	0.043**	-0.017	-0.003	0.034	0.045*
	[0.356]	[0.558]	[1.346]	[2.384]	[-0.830]	[-0.146]	[1.262]	[1.711]
Marital Status	0.061***	0.059***	0.037**	0.023	0.012	0.013	0.030	0.029
	[4.817]	[4.348]	[2.056]	[1.231]	[0.505]	[0.627]	[1.312]	[1.309]
Urban	0.034***	0.047***	0.018	0.048***	0.049**	0.076***	0.033*	0.050**
	[3.134]	[4.329]	[1.109]	[2.956]	[2.460]	[3.845]	[1.656]	[2.383]
Province Dummies		Y		Y		Y		Y
Observations	1,001	1,001	832	832	764	764	686	686
R-squared	0.164	0.233	0.176	0.291	0.128	0.294	0.187	0.223

2.5.7.2 Pooled cross-sectional analysis

Cross-sectional results presented in the previous section provide suggestive evidence that it is precisely those high-educated households that were exposed to famine early on in their lives that have higher dietary diversity scores. In this subsection, I extend this analysis by pooling the four waves {2004, 2006, 2009 and 2011} together. In addition, I employ the extended pooled sample that also incorporates the 1997 and 2001 wave. The results of estimating Equation (15) and (15.1) on the pooled cross-sectional sample are shown in *Table 1.16.*

Table 1. 16: Effects of interaction of Famine inverted term (EDR_inv [%/diff]) and Education on Dietary Diversity Score; Pooled regressions (1997-2011) and (2004-2011).

Variables are defined as described in Table 1.2. EDR_inv [%] represents inverted value of Excess Death Rate, calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958);); EDR_inv [diff] represents inverted value of Excess Death Rate calculated as difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using 20 food groups.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets. All regressions include wave dummies, while regressions (2), (4), (6) and (8) include both wave and province dummies.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1997-2011				2004-2011			
log (HH total income)	0.061***	0.051***	0.060***	0.051***	0.060***	0.050***	0.060***	0.050***
	[18.894]	[16.390]	[18.748]	[16.392]	[15.242]	[13.122]	[15.182]	[13.121]
Education	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000
	[1.418]	[0.254]	[0.380]	[0.078]	[0.826]	[0.370]	[0.006]	[0.198]
EDR_inv [%]	0.034***				0.087***			
	[3.303]				[6.337]			
Age	0.028*	0.002	0.029*	0.002	0.007	-0.029	0.011	-0.029
	[1.802]	[0.155]	[1.875]	[0.130]	[0.323]	[-1.294]	[0.456]	[-1.312]
EDR_inv[%] x Education	0.002*	0.004**			0.002	0.002		
	[1.885]	[2.521]			[0.968]	[1.307]		
Gender	-0.005	-0.007	-0.005	-0.007	0.001	-0.006	0.001	-0.006
	[-0.921]	[-1.461]	[-0.934]	[-1.466]	[0.116]	[-0.903]	[0.114]	[-0.897]
HH size	-0.007***	-0.002**	-0.006***	-0.002**	-0.006***	-0.000	-0.006***	-0.000
	[-6.975]	[-2.472]	[-6.481]	[-2.481]	[-4.440]	[-0.202]	[-3.981]	[-0.209]
Han nationality	-0.008*	0.007*	-0.010***	0.007*	-0.017***	0.003	-0.016***	0.003
	[-1.943]	[1.705]	[-2.590]	[1.708]	[-3.310]	[0.516]	[-3.299]	[0.518]
Old home food habits	0.004	0.014*	0.005	0.014*	0.009	0.019*	0.012	0.019*
	[0.484]	[1.930]	[0.747]	[1.942]	[0.819]	[1.941]	[1.119]	[1.949]
Marital Status	0.035***	0.032***	0.034***	0.032***	0.037***	0.033***	0.036***	0.033***
	[5.561]	[5.338]	[5.481]	[5.355]	[3.979]	[3.702]	[3.883]	[3.713]
Urban	0.028***	0.041***	0.028***	0.041***	0.032***	0.054***	0.034***	0.054***
	[5.617]	[8.284]	[5.786]	[8.291]	[4.077]	[6.787]	[4.261]	[6.794]
EDR_inv [diff]			0.035***				0.082***	
			[3.632]				[6.422]	
EDR_inv [diff] x Education			0.003**	0.004***			0.003*	0.003
			[2.438]	[2.697]			[1.692]	[1.490]
Wave Dummies	Y	Y	Y	Y	Y	Y	Y	Y
Province Dummies		Y		Y		Y		Y
Observations	5,748	5,748	5,748	5,748	3,283	3,283	3,283	3,283
R-squared	0.180	0.247	0.185	0.247	0.193	0.271	0.199	0.271
AIC	-10,225	-10,700	-10,256	-10,701	-5,476	-5,797	-5,504	-5,798
BIC	-10,112	-10,547	-10,143	-10,548	-5,384	-5,669	-5,412	-5,670

Columns (1)-(4) show the results for the 1997-2011 pooled sample, while Columns (5)-(8) show the results for the 2004-2011 pooled sample. As before, all specifications include wave dummies, while even specifications also include province dummies. Estimated coefficient on the interaction term *EDR_inv x Compl. Years Education (Head)* is positive and significant at 10% (5%) level in both Column (1) and (2), 0.002 (0.04 food groups) and 0.004 (0.08 food groups) respectively, suggesting that controlling for wave and province unobserved heterogeneity, households whose heads had higher number of years of completed education and those with less prior exposure to famine are those with higher dietary diversity scores. Similarly, estimated coefficient on the interaction term *EDR_inv[diff] x Compl. Years Education (Head)* is positive and significant at 5% (1%) level in both Column (3) and (4), 0.003 and 0.004 respectively, suggesting that more-educated (heads of) households with less prior exposure to famine are precisely those that have higher dietary diversity scores, irrespective of the chosen proxy for prior famine exposure. Results shown in Columns (5)-(8) further corroborate these findings.

2.5.7.3 Pooled split sample analysis

Here, I split the sample into those who were born before and during the Great famine (1954 - 1962) and those born after the famine (1963 – 1988), for the same reasons as elaborated in 6.6.3 sub-section. Columns (1)-(4) show the results for the famine cohort (1954-1962) of the pooled sample, while Columns (5)-(8) show the results for the post-famine cohort (1963-1988) of the pooled sample. Odd columns include wave dummies only, while even columns include province dummies too.

Variable of interest is the interaction term between EDR and education. As per *Table 1.17*, in case of *EDR [%] x Education*, this coefficient is positive, but it is not significant at conventional levels neither in famine generation, nor in post-famine generation. In case in the interaction term *EDR [diff] x Income,* the coefficient is significant at 10% level in famine generation and the effect disappears in post-famine generation. Interestingly, the significance also disappears in case of *EDR [%] and EDR [diff],* when comparing famine and post-famine cohorts.

These results suggest that the heterogenous effect of famine on DDS by education level diminishes over generations.

Table 1. 17: Effects of interaction of Famine inverted term (EDR_inv [%/diff]) and Education on Dietary Diversity Score; Pooled regression (2004-2011), split sample Cohort 1954-1962 and Cohort 1963-1988

Variables are defined as described in Table 1.2. EDR_inv [%] represents inverted value of Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR_inv [diff] represents inverted value of Excess Death Rate calculated as difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using 20 food groups.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets. All regressions include wave dummies, while regressions (2), (4), (6) and (8) include both wave and province dummies.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Born 1954-1962				Born 1963-1988			
log (HH total income)	0.054***	0.045***	0.054***	0.045***	0.032**	0.033***	0.032**	0.033***
	[7.338]	[6.149]	[7.382]	[6.110]	[2.586]	[2.826]	[2.589]	[2.809]
Education	-0.001	-0.001	-0.003	-0.002	-0.002	0.002	0.007	0.004
	[-0.305]	[-0.268]	[-1.189]	[-0.832]	[-0.185]	[0.197]	[0.785]	[0.446]
EDR_inv [%]	0.081**				0.057			
	[2.336]				[0.420]			
Age (Head)	0.156	0.182	0.170	0.185	-0.987**	-0.847**	-1.051***	-0.852**
	[1.049]	[1.318]	[1.153]	[1.340]	[-2.551]	[-2.294]	[-2.715]	[-2.288]
EDR_inv [%] x Education	0.004	0.005			0.006	0.003		
	[0.928]	[1.107]			[0.380]	[0.222]		
Gender	0.056***	0.034***	0.054***	0.034***	-0.007	-0.013	-0.001	-0.012
	[3.668]	[2.686]	[3.635]	[2.679]	[-0.272]	[-0.458]	[-0.025]	[-0.411]
HH size	-0.014***	-0.004	-0.013***	-0.004	-0.009**	-0.005	-0.008**	-0.005
	[-4.425]	[-1.103]	[-4.134]	[-1.119]	[-2.148]	[-1.252]	[-2.103]	[-1.255]
Han nationality	-0.012	0.015	-0.008	0.014	-0.018	0.002	-0.017	0.002
	[-1.237]	[1.292]	[-0.835]	[1.220]	[-1.077]	[0.095]	[-1.054]	[0.105]
Old home food habits	-0.038*	-0.019	-0.033	-0.019	-0.090	-0.068	-0.090	-0.068
	[-1.753]	[-0.970]	[-1.570]	[-0.939]	[-1.505]	[-1.358]	[-1.519]	[-1.357]
Marital Status	0.046**	0.006	0.044**	0.007	0.016	0.005	0.020	0.005
	[2.142]	[0.282]	[2.056]	[0.327]	[0.798]	[0.250]	[1.001]	[0.260]
Urban	0.037**	0.067***	0.039***	0.067***	0.008	0.051*	0.011	0.052*
	[2.488]	[4.382]	[2.649]	[4.410]	[0.360]	[1.908]	[0.521]	[1.936]
EDR_inv[diff]			0.059*				0.167	
			[1.828]				[1.468]	
EDR_inv[diff] x Education			0.007*	0.007*			-0.006	-0.000
			[1.749]	[1.675]			[-0.488]	[-0.019]
Wave dummies	Y	Y	Y	Y	Y	Y	Y	Y
Province dummies		Y		Y		Y		Y
Observations	918	918	918	918	360	360	360	360
R-squared	0.210	0.303	0.218	0.304	0.203	0.321	0.217	0.321
AIC	-1,528	-1,632	-1,538	-1,633	-630	-676	-636	-675
BIC	-1,456	-1,531	-1,465	-1,532	-572	-594	-578	-594

2.5.8 Early life famine and dietary diversity score: heterogenous effect by gender

2.5.8.1 Cross-sectional analysis

In this subsection I test whether the effect of childhood famine on dietary diversity score later in life differed by gender, that is, if the famine had a differential effect on dietary diversity for males and females. In particular, I estimate Equation (16) on each of the four cross-sections {2004, 2006, 2009 and 2011}. Results are shown in *Table 1.18*.

Table 1. 18: Effects of interaction of Famine inverted term (EDR_inv [%]) and Gender on Dietary Diversity Score; Cross sectional regressions

Variables are defined as described in Table 1.2. EDR_inv [%] represents inverted value of Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); Dependent variable Dietary Diversity Score is calculated using 20 food groups.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets.*

	Dietary Diversity Score			
	2004	2006	2009	2011
log (HH total income)	0.076***	0.065***	0.042***	0.063***
	[9.487]	[7.335]	[5.679]	[8.585]
Education	0.001	0.003***	0.002*	0.001
	[1.132]	[2.631]	[1.715]	[0.397]
EDR_inv [%]	0.070*	0.023	0.042	0.108**
	[1.920]	[0.559]	[0.714]	[2.123]
Age	0.010	0.092**	0.009	-0.112**
	[0.272]	[2.001]	[0.161]	[-1.982]
EDR_inv [%] x Gender	0.007	0.083*	0.074*	-0.004
	[0.197]	[1.924]	[1.789]	[-0.068]
Gender	-0.006	-0.048**	-0.016	0.004
	[-0.330]	[-2.186]	[-0.604]	[0.173]
HH size	-0.009***	-0.009***	-0.005	-0.002
	[-3.722]	[-2.886]	[-1.580]	[-0.460]
Han nationality	-0.014	0.003	-0.022**	-0.038***
	[-1.536]	[0.291]	[-2.085]	[-3.231]
Old home food habits	0.006	0.023	-0.018	0.034
	[0.340]	[1.258]	[-0.855]	[1.256]
Marital Status	0.061***	0.045**	0.018	0.029
	[4.719]	[2.337]	[0.736]	[1.270]
Urban	0.034***	0.019	0.050**	0.033
	[3.087]	[1.153]	[2.479]	[1.635]
Observations	1,001	832	764	686
R-squared	0.163	0.178	0.129	0.187

As before, the estimated coefficient on the inverse normalized excess death rate measure *EDR_inv* is positive and significant at 5% (10%) level 2004 and 2011, suggesting that households who had lower prior exposure to famine have higher dietary diversity scores later in life. The estimated coefficient on the *EDR_inv x Gender (Head)* interaction term is mostly positive, and statistically significant at 10% level in 2006 and 2009. In Annex 1, *Table A.1.14* I repeat this cross-sectional analysis by using *EDR_inv[diff]* which captures the inverse of the normalized excess death rate *EDR[diff]* between time period 1959-1961 and 1956-1958 in province *j*, where the household is located. Results remain quantitatively and qualitatively similar.

This result provides weak evidence that among male individuals, higher dietary diversity score is driven by those who had lower exposure to famine in their childhood, or were located in the areas less affected by the famine. We further test this conjecture in the pooled regression setting in the next subsection.

2.5.8.2 Pooled cross-sectional analysis

Cross-sectional results presented in the previous section provide suggestive evidence that it is precisely those male individuals that were less exposed to famine early on in their lives that have higher dietary diversity scores. In this subsection, I extend this analysis by pooling the four waves {2004, 2006, 2009 and 2011} together. In addition, I employ the extended pooled sample that also incorporates the 1997 and 2001 wave. The results of estimating Equation (17) on the pooled cross-sectional sample are shown in *Table 1.19*.

Table 1. 19: Effects of interaction of Famine inverted term (EDR_inv [%/diff]) and Gender on Dietary Diversity Score; Pooled regressions (1997-2011) and (2004-2011).

Variables are defined as described in Table 1.2. EDR_inv [%] represents inverted value of Excess Death Rate, calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958);); EDR_inv [diff] represents inverted value of Excess Death Rate calculated as difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using 20 food groups.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include wave dummies.*

	(1)	(2)	(3)	(4)
	1997-2011		2004-2011	
log (HH total income)	0.061***	0.060***	0.060***	0.060***
	[18.885]	[18.759]	[15.242]	[15.211]
Education	0.002***	0.002***	0.002***	0.002***
	[5.063]	[4.798]	[2.953]	[2.730]
EDR_inv [%]	0.023		0.058**	
	[1.301]		[2.490]	
Age	0.027*	0.029*	0.007	0.010
	[1.792]	[1.869]	[0.303]	[0.450]
EDR_inv [%] x Gender	0.024*		0.042*	
	[1.747]		[1.751]	
Gender	-0.016*	-0.019**	-0.018	-0.020*
	[-1.791]	[-2.134]	[-1.594]	[-1.777]
HH size	-0.007***	-0.006***	-0.006***	-0.006***
	[-6.979]	[-6.500]	[-4.406]	[-3.944]
Han nationality	-0.008*	-0.010***	-0.017***	-0.016***
	[-1.938]	[-2.654]	[-3.267]	[-3.304]
Old home food habits	0.003	0.005	0.008	0.011
	[0.475]	[0.709]	[0.786]	[1.067]
Marital Status	0.036***	0.036***	0.040***	0.039***
	[5.624]	[5.557]	[4.163]	[4.061]
Urban	0.028***	0.028***	0.032***	0.034***
	[5.601]	[5.742]	[4.088]	[4.251]
EDR_inv[diff]		0.027		0.059***
		[1.643]		[2.767]
EDR_inv[diff] x Gender		0.028*		0.042*
		[1.699]		[1.921]
Wave Dummies	Y	Y	Y	Y
Observations	5,748	5,748	3,283	3,283
R-squared	0.180	0.184	0.193	0.199
AIC	-10,224	-10,695	-5,252	-5,695
BIC	-10,111	-10,542	-5,139	-5,542

Columns (1)-(2) show the results for the 1997-2011 pooled sample, while Columns (3)-(4) show the results for the 2004-2011 pooled sample. As before, all specifications include wave dummies, however in this case I cannot include a vector of province dummies since those would be perfectly colinear with the interaction term *EDR_inv x Gender (Head),* which in that case could not be estimated. Estimated coefficient on the interaction term *EDR_inv x Gender (Head)* is positive and significant at 10% level in both Column (1) and (3), suggesting that controlling for wave unobserved heterogeneity, among households with less prior exposure to famine, those whose head is male tend to have higher dietary diversity scores. Similarly, estimated coefficient on the interaction term *EDR_inv[diff] x Gender (Head)* is positive and significant at 10% level in both Column (2) and (4), suggesting that among households with less prior exposure to famine, those whose head is male tend to have higher dietary diversity scores, irrespective of the chosen proxy for prior famine exposure. Having said that, the magnitude of the coefficient suggests that the difference in number of consumed food groups between male and female is 0.5

2.5.8.3 Pooled split sample

I perform pooled split sample analysis of famine and post-famine cohorts, with respect to *EDR_inv* [%] *x Gender (Head)* and *EDR_inv* [diff] *x Gender (Head).* This analysis did not reveal any differences between famine and post-famine cohorts in heterogenous effects of famine on DDS by gender. The results are reported in Annex 1, *Table A.1.15.*

2.5.9 Extending the data sample: inclusion of the 1997 and 2000 waves

Most of the existing literature using the CHNS dataset, focuses their analysis on one or two waves (cross-sections). To the best of my knowledge, this is one of very few studies that attempts to collate data from several different waves, in an attempt to create a substantially larger data sample, which is not prone to selection issues. In the analysis so far, I mainly focused on analysing four waves: 2004, 2006, 2009 and 2011, since these are the waves for which the definition of food groups, which are used for the computation of dietary diversity scores, is consistent. In this subsection, I augment my data sample to include the 1997 and 2000 wave. In order to arrive at a consistent measure of dietary diversity score, I merged food composition tables (CFCT - 1991) used for waves 1997 and 2000 with food composition tables (CFCT – 2002/04) used in 2004, 2006, 2009, 2011. To do that, I first had to translate CFCT-1991 to CFCT 2002/04 and that process is graphically presented in Annex 1 *Tables A.1.2 and A.1.3.* Sub-section "*Methodology*" within section 2.4, contains description of this process.

I then proceed to re-estimate all of my analysis on this augmented sample. Appendix 1 shows the results. *Table A.1.13* shows the results of the pooled regressions, including six waves {1997, 2000, 2004, 2006, 2009 and 2011}. As reported in *Table A.1.13*, the estimated coefficient on EDR [%] and EDR[diff] are negative and statistically significant at 1% level. They are also of the very similar magnitude as those reported for the smaller sample (*Table 1.4*). The estimated coefficient on *Age* is positive and significant at 10% level in Columns 1 and 2, for linear specifications. In Columns 3 and 4, when I introduce the quadratic age term *Age^2*, the estimated coefficient on *Age* is positive and significant at 1% level. This suggests that older individuals have higher dietary diversity scores. However, given that the estimates on the quadratic age term *Age^2*, are negative and statistically significant at 1% level in both Column 3 and 4, suggests that this relationship is not linear, but rather concave.

In *Table A.1.14* I further analyse this result, by running pooled regressions on the augmented sample, which now include the interaction terms *EDR [%] x Age (Head)* and *EDR [%] x Age^2 (Head),* as specified in Equation (7) and (9). Columns 1-4 include wave dummies, while columns 5-8 include wave and province dummies. As we can see from Column (1) and Column (5), the estimated coefficient on the interaction terms *EDR x Age (Head)* is negative and significant at 1% (10%) level respectively. As for the magnitude, these coefficients translate to 2.68 food groups with wave dummy only and 1.36 food groups after including both wave and province dummy. Similarly, ss we can see from Column (2) and Column (6), the estimated coefficient on the interaction terms *EDR [diff] x Age (Head)* is negative and significant at 1% (10%) level respectively. This suggests that older individuals, who were directly or indirectly exposed to the famine of 1959-61, tend to have lower dietary diversity scores later in life.

By inspecting the estimated coefficients on the interaction terms *EDR [%] x Age^2 (Head)* and EDR *[diff] x Age^2 (Head)* in columns (3) and (4), we see that they are negative and significant at 5% (10%) level respectively. As for the magnitude, this corresponds to 1.76 food groups in case of EDR [%] and 0.1 food groups when EDR is measured as a difference between the famine and pre-famine period. This suggests that there is a concavity in this relationship, suggesting that the oldest of the individuals that were exposed to famine have the lowest dietary diversity scores later in life. When both wave and province dummies are included, as shown in Columns (7) and (8), the estimated coefficients on these interaction terms are still negative, but to longer statistically significant. Note that the results presented using the augmented sample are of different sign relative to those presented in Table 1.5. While in Table 1.5, which covers the sample between 2004 and 2011, the estimated coefficients on the interaction term *EDR [%] x Age (Head)* and *EDR [%] x Age^2 (Head)* are positive, in Table A.1.14 (which covers wave 1997 to 2011) they are negative. This can potentially be driven by attrition rate, in that some of the respondents who were alive in 1997 and 2000 waves, were no longer among the living in later waves. This suggests that results presented in Table A.1.14 are mainly driven by the old respondents, born before the great famine, who were still captured by the 1997 and 2000 survey waves.

In *Tables A.1.15 to A.1.18* I repeated prior analysis using six food groups for the computation of dietary diversity score (*Table A..1.15 and Table A.1.16*) and using twelve food groups for the computation of dietary diversity score (*Table A.1.17 and Table A.1.18*). My results remain unchanged.

In *Table A..1.19* I repeat the birth-year analysis as described in Section 2.5.5 on this augmented sample. As before, we see that most birth groups around the Great Chinese Famine were affected negatively in terms of their dietary diversity scored later in life, as evidence by negative coefficients on the interaction terms. However, it seems that those born in 1957 were more affected by the others, as evidenced by the magnitude of the estimated coefficient (-0.068 and -0.004, for EDR and EDR [diff] as proxies for famine exposure). The two coefficients correspond to 1.36 and 0.08 food groups. Moreover, these are the only coefficient estimates that are statistically significant (at 5% and 10% level, respectively).

Based on information criteria tests (AIC and BIC) specifications which include both province and wave dummy variables seem to exhibit better relative goodness-of-fit comparing to those that include only wave dummies. This holds irrespective of number of food groups being used in the model (6, 12 or 20).

Additionally, AIC and BIC tests show that models with higher number of observations exhibit better relative goodness-of-fit comparing to those with lower number of observations. This is the case when comparing cross-sectional estimations (2004, 2006, 2009, 2011), cross-sectional with repeated cross-sectional (2004, 2006, 2009, 2011, 2004-2011) or two sets of repeated cross-sectional (1997-2011,2004-2011).

Based on information criteria tests (AIC and BIC) specifications which include both province and wave dummy variables seem to exhibit better relative goodness-of-fit comparing to those that include only wave dummies. This holds irrespective of number of food groups being used in the model (6, 12 or 20).

Additionally, AIC and BIC tests show that models with higher number of observations exhibit better relative goodness-of-fit comparing to those with lower number of observations. This is the case when comparing cross-sectional estimations (2004, 2006, 2009, 2011), cross-sectional with repeated cross-sectional (2004, 2006, 2009, 2011, 2004-2011) or two sets of repeated cross-sectional (1997-2011,2004-2011).

3 Long-term Effects of Famine on Macronutrient Composition in Diet

Episodes of famine were very common and severe in the past, and in 20th century alone famines claimed more than 70 million lives (Devereux, 2000). However as per *Section One* of this dissertation it appears that they are not matter of the past but will continue to repeat in the future. Famines are rarely isolated incidents, and they are often connected with conflicts, extreme climate events, either combined or separately (Oberg et al., 2021; Slavin, 2016). The previous section, *Long-term effects of famine on Dietary Diversity Score*, shows that exposure to famine can affect what famine survivors consume decades after the adverse event. Specifically, those who experienced famine consume less diverse diet, relative to those who have not experienced famine, and this effect is proportional to famine severity – the more intense the famine was, the less diverse the diet will be. In addition to this, the previous section shows that there is a particular age-cohort which is most affected by famine, and this is measured by dietary diversity score. This age cohort include those who were 2-5 years old in time of famine. In addition to those who were exposed to famine, the results from the previous section suggest that diet of those who were born after the famine, yet in the provinces affected by famine also consume less diverse diet.

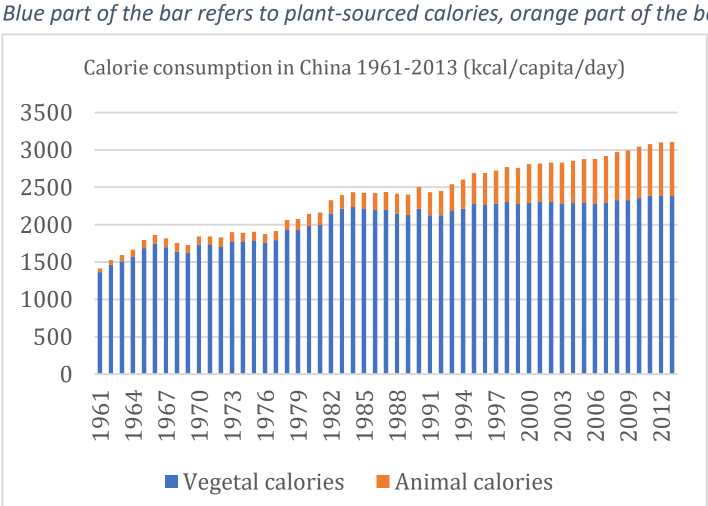
Variable 'old home' food habits, which is relevant not only in *Section Two*, but remains relevant throughout this dissertation offered inconclusive results. Having said that, 'old home' food habits, which present informant's continuation of eating patterns which they had in their childhood revealed positive correlation, only when I included province dummies in pooled regressions. In addition to this 96% of the respondents from CHNS dataset claimed that they keep 'old home' food habits, and it comes as surprise that estimation coefficients do not show higher statistical significance.

In addition to this, the previous section showed that DDS of those affected by famine is simply lower than of individuals not affected. However, this result does not tell us the composition of the DDS of those who experienced famine. Hence, it remains unclear which nutrients were deficient in diets with lower DDS. Hence, this section analyses the missing component, and reveals an additional dimension of long-term effects of famine, proxied by the composition of nutrients in diet.

3.1 Evidence from the Existing Literature

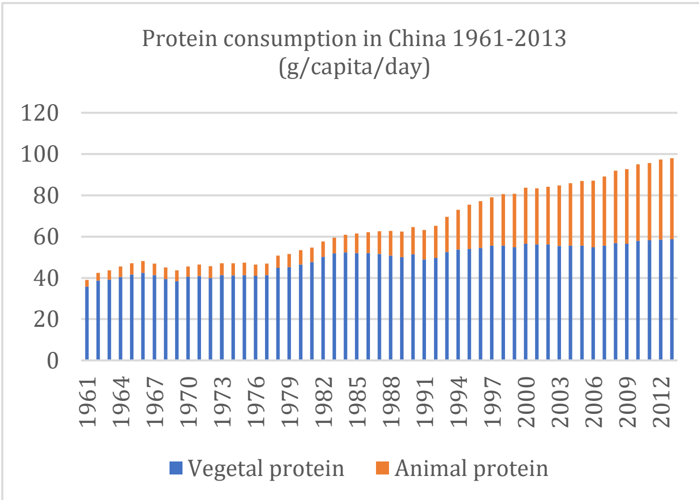
The concept of Nutrition Transition, first described by Popkin (1993) posits that while a society is progressing through the stages of the nutrition transition, increasing income of the population is coupled with changes in energy balance. On one hand, energy dense foods, which become more dominant in later stages of nutrition transition, drive increase in human energy intake. On the other hand, human energy expenditure decreases over the course of nutrition transition, as larger share of the population replaces relatively active lifestyle to predominantly sedentary one. Concurrently, the composition of diet changes from less diverse and cereal-dominated one, to diets with more diverse foods where animal sourced foods, as well as fruits and vegetables increase their relative share. From macronutrient perspective, decline in cereal consumption leads to decline in carbohydrates intake, while increase in the consumption of animal products lead to protein and fat consumption. Graphs 2.1-2.4 which illustrate this food and macronutrient transition are captured by FAO Food Balance Sheets (FBS). As per the graphs, in the period 1961-2013, daily calorie intake in China increased from 1415 to 3109 kcal (120% increase), protein intake increased from 39 to 98g (150% increase) while fat intake increased from 14.5 to 95g in the given period (550% increase). This evidence is in line with nutrition transition concept. In addition to this, the graphs show that while both plant and animal sourced nutrients have increased between 1961 and 2013, increase of animal sourced nutrients and calories were occurring at much higher rate.

Graph 2. 1: Trend in average total calorie consumption in China in period 1961-2013 (kcal/capita/day)
Blue part of the bar refers to plant-sourced calories, orange part of the bar refers to animal-sourced calories



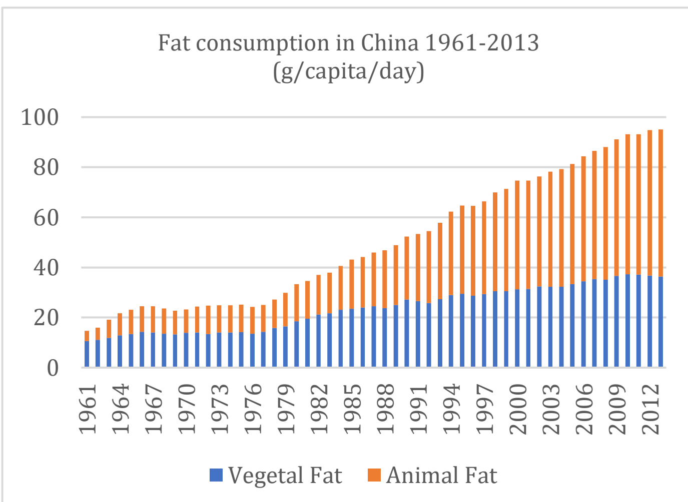
Source: FAOstat

Graph 2. 2: Trend in average protein consumption in China in period 1961-2013 (g/capita/day)
Blue part of the bar refers to plant-sourced protein, orange part of the bar refers to animal-sourced protein



Source: FAOstat

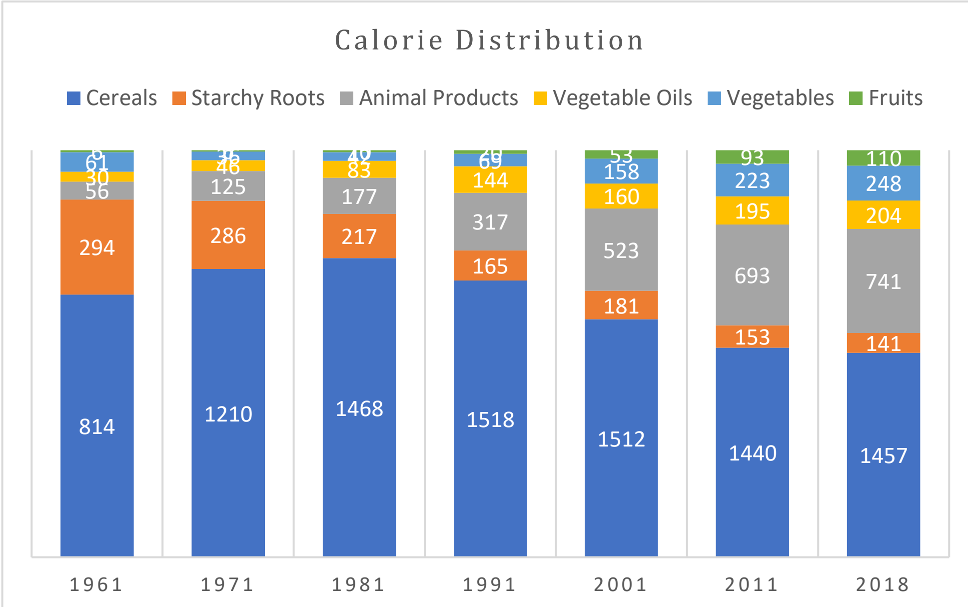
Graph 2. 3: Trend in average fat consumption in China in period 1961-2013 (g/capita/day)
Blue part of the bar refers to plant-sourced fat, orange part of the bar refers to animal-sourced fat



Source: FAOstat

Graph 2. 4: Share of sources of calories (%) and absolute values of sources of calories (kcal) in Chinese diet in period 1961-2018

Dark blue part of the bar represents cereals, orange represents starchy roots, gray represents animal products, yellow represents vegetable oils, light blue represents vegetables and green part of the bar represents fruits.



Source: FAOstat

FBS include numerous elements which could be used to monitor nutrition transition in a country. However, the dataset provides only national average, and does not provide data differentiated by gender, age, nationality, region in China etc. In that respect CHNS provide much deeper insight into diets in China. As per Zhai et al. (2014), who use CHNS in their analysis, share of energy from fat increased from 21.8% in 1991, to 32% in 2011. In the same period, share of energy from carbohydrates decreased from 66% to 54.3% and share of energy from protein increased from 11.8% to 13.3% (Zhai et al., 2014). The same source finds that calories from animal sourced food increased from 240 to 333 in the period 1991-2011, and this refers to the population between 19-59 years old. In the same population and the same period, calories from coarse grains decreased from 114 to 40. These trends coincide with the ones from FAO FBS.

It is important to note that while calories derived from animal products increased from 177 to 741 in the period 1981-2018, the calories from cereals decreased from 1468 to 1457 in the same period (Graph 2.4). Similarly, Pingali (2007) describes westernization of Asian diets from traditional, rice-dominated diets to the ones dominated by increased share of wheat, fruits, vegetables as well as high protein and energy dense foods. Thus, the results from the previous Section where I show that DDS of those affected by famine is lower than of those who were not affected, framed in the Nutrition Transition theory suggest that those who were directly or indirectly exposed to famine are in earlier stage of nutrition transition comparing to those who were not exposed to famine.

In addition to share of cereals in diet, there are two more metrics used in the literature, which explores determinants of dietary patterns. These metrics are staple diversity score (SDS) and the proportion of coarse staple consumption (PoCS) (Chang et al., 2018). As per authors, SDS represents variety of staple foods in one's diet, while PoCS represents share of coarse staples in total staples. As for the coarse staple consumption, the authors found that share of coarse grains in total consumption is driven not only by change in cropping structure, but also by increase in purchase power.

Literature, which analyse determinants of micro- and macronutrients intake and food consumption, primarily focus on socio-economic determinants. Education, income, marital status, employment, smoking habits, food access and social norms are some of those determinants (Haste et al., 1990; Parpia, 1995; You et al., 2016). To the best of my knowledge this dissertation presents the first study which examines the relationship between exposure to famine and macronutrient intake.

3.2 Hypothesis development

In *Section Two*, I found that exposure to famine, either through personal experience, or though living in areas affected by famine has a negative impact on dietary diversity score (DDS), and to the best of my knowledge that is the first analysis which established that relationship. The following step in my analysis is to determine how lower DDS translates to nutritional intake. The existing literature does not provide sufficient evidence on how exposure to famine affects nutrient intake. This chapter aims to fill this gap.

As described earlier the main goal of this dissertation is to study the effects of famine on nutrient intake later in life. In *Section Two*, I use DDS to explore this. In *Section Three* I use share of carbohydrates, fat and protein in total diet. In *Section Four*, I expand my analysis by studying effects of famine on lifestyle choices such as food expenditures, alcohol and tobacco intake and others.

The existing evidence from nutrition transition literature suggests that as a society moves from less to more affluent status, consumption patterns change from cereal-based less diverse, to more diverse dietary patterns. Additionally, *Section Two* of this dissertation shows that exposure to famine has negative effects on DDS.

Literature analysed in *Section Two* shows that the Great Famine survivors tend to waste less food (Ding et al., 2022), have higher bank savings (Chen et al., 2018) and express more conservative and less risky behaviour (Hu et al., 2017; Zhang, 2017) relative to those who were not affected by the famine. In other words, they exhibit more frugal behaviour. Hence, I hypothesize that those who were more exposed to the Great Famine eat more cereals-based diet, as that staple can provide sufficient caloric intake, yet it is less perishable relative to fruits, vegetables and meat. Additionally, it is cheaper, and generates less food waste. In nutritive terms, this means that share of carbohydrates in total diet of those who were affected by famine is higher and share of fat and protein in total diet is lower than of those who were not affected.

Using nutrition transition framework, those who were affected by famine are in an earlier stage of nutrition transition relative to those who were not affected by famine.

H1: Share of carbohydrates in diet is higher for famine survivors who were more affected by the famine, relative to those who were less affected.

H2: Share of proteins in diet is lower for famine survivors who were more affected by the famine, relative to those who were less affected.

H3: Share of fat in diet is lower for famine survivors who were more affected by the famine, relative to those who were less affected.

3.3 Data

Datasets used in this chapter are the same as used in *Chapter One*. Those are China Health and Nutrition Survey (CHNS), and mortality data used Meng et al. (2015), and which are sourced from China Statistics Bureau data.

As described earlier CHNS dataset is jointly developed and implemented by UNCH-CPC and CCDCP to capture the impacts of economic and social transformations of Chinese society in post-trade liberalization period on health and nutrition of the population. To that end, a set of household and individual socio-economic and demographic factors have been collected. Detailed description of the sampling methodology, geographical and temporal coverage and other information about the dataset can be found in "*Data*" section of *Section Two*.

Health and Nutrition Survey subsection of "*Data"* section in *Section Two* describes how dietary data have been collected in CHNS. To obtain information about nutritive characteristics of the foods consumed by survey participants, China Food Composition tables were combined with food records obtained in CHNS. Consumption of macronutrients is being calculated per person, by dividing the total amount of calories and macronutrients consumed by number of household members. In my specifications, per person refers to head of household, so that interpretation of the result is consistent with other Sections of dissertation.

China Food Composition Tables (CFCT) is a document which contains information about:

- Nutrient content of foods
- Amino acid content of foods
- Fatty acid content of foods
- Choline, Biotin, Pantothenic acid, Vitamin K and Vitamin D content of foods

Nutrient content of food is particularly important for this dissertation, as it contains carbohydrate, protein and fat content, as well as caloric value of each food item collected through CHNS.

In CFCT, the *energy value* (kcal) of a food item is calculated as the sum of all energy producing nutrients, and those are carbohydrates, protein, fat and alcohol; *protein content* (g) is calculated by multiplying total nitrogen with corresponding protein conversion factor; *carbohydrate content* (g) was calculated by applying a difference subtracting method; and total *fat content (g)* represents the total content of crude fat (China CDC, 2004).

Values of protein, fat and carbohydrates, as well as of kcal in CHNS present 3-day average values per person, and summary statistics present absolute intake, and relative share in total intake of the three macronutrients.

Table 2. 1: Summary Statistics of key dependent variables
Share of fat presents share of fat in total diet of household head per day (%); Share of protein presents share of protein in total diet of household head per day (%); Share of carbohydrates presents share of carbohydrates in total diet of household head per day (%); Protein intake presents absolute intake of protein of household head per day (g/capita/day); Fat intake presents absolute intake of fat of household head per day (g/capita/day); Carbohydrates intake presents absolute intake of carbohydrates of household head per day (g/capita/day); kcal intake presents absolute intake of calories of household head per day (kcal/capita/day).

	mean	sd	min	max	N
Share of fat	0.15	0.07	0.01	0.55	3,283
Share of protein	0.14	0.03	0.04	0.37	3,283
Share of carbohydrates	0.71	0.09	0.32	0.91	3,283
Protein intake	71.27	27.04	11.59	441.91	3,283
Fat intake	71.95	38.76	1.78	439.10	3,283
Carbohydrates intake	360.69	125.57	41.84	979.16	3,283
kcal intake	2449.12	744.76	591.99	6084.41	3,283

As per table above, average share of fat and protein in total diet is around 15% each, while relative share of carbohydrates in diet is around 70%.

Comparing to FAO Food Balance Sheets, absolute intake of protein and fat is slightly lower. It is 71 g/day for both fat and protein in CHNS sample, while FBS shows that the values are 80g/day for fat and 85 g/day for protein. Average intake of carbohydrates is 360 g/capita/day, while daily caloric intake is 2449 kcal/capita/day. Daily caloric intake per FBS was 2797 kcal/capita/day in 2011.

3.4 Methodology and Empirical Strategy

Methodology

In *Section Two* of the dissertation, I find that famine survivors have dietary patterns with lower DDS comparing to those individuals, who were not affected by famine, and that famine severity is inversely correlated with DDS. Additionally, I found that age at the time of famine does matter, and that those who were 2-5 years old during the famine were affected more than other age cohorts. However, *Section Two* did not reveal how is DDS of famine survivors different from those who have not experienced famine. To that end, I introduce new dependent variables in *Section Three,* and these are: Calories, Carbohydrates, Protein, Fat, Share of carbohydrates/protein/fat in diet. The new variables allow me to analyse whether nutritive composition of diets of those exposed to famine differs from those who were not exposed.

<u>Calories, Carbohydrates, Protein, Fat, Share of carbohydrates/Protein/Fat in diet</u>

Calories are being measured by number of calories (kcal); ***Carbohydrates, Proteins*** and ***Fat*** are being measured in grams; ***Share of carbohydrates/protein/fat in diet*** are being measured by share of one of the three nutrients (g) in sum of all three nutrients (g). As described in "*Data*" section of this chapter, the information about grams of the nutrients is being created by combining household consumption data from CHNS with nutrient content of each food product from CFCT.

$$Nutrient\_share_{n,head,j} = \frac{Nutrient_n}{\sum_{n=1}^{3} Nutrient_n}$$

Where *n* depicts one of the three nutrients (carbohydrates, fat or protein), *head* denotes household head in province *j*.

<u>Famine</u>

Similar to approach in *Section Two*, EDR [%] measures the **ratio** between average value of death rate during the three years of famine, and average value of death rate during the three years prior to famine.

$$EDR[\%] = \frac{Avg\ Famine\ (1959 - 1961)}{Avg\ Famine\ (1956 - 1958)}$$

Unlike in *Section Two* where I used both **EDR ratio** and **EDR difference**, in *Section Three* I will use only **EDR ratio** as a proxy for famine, since the results obtained through both metrics were consistent in *Section Two*.

Empirical Strategy

3.4.1 The effect of early life famine on macro-nutrient consumption later in life

3.4.1.1 Determinants of macro-nutrient consumption

One of the main goals of this study is to assess the effect of early life exposure to famine on macro-nutrient consumption later in life. Existing literature (Li & An, 2015; Li et al., 2010, 2011; Shi et al., 2013) has mostly focused on using select waves from the CHNS dataset in their analysis. As argued in *Section Two*, a potential concern with this approach is that it gives rise to selection issues, in that drawing inference from a single study wave can have certain limitations when it comes to external validity of the obtained results. While in my approach above I use four different waves to conduct my analysis, a potential issue with the approach above is that it does not take into account the fact that certain time-specific factors prevalent in each wave could be driving the obtained results. To address this issue, I pool the four cross-sections together and estimate a pooled cross-sectional regression on a pooled sample consisting of four waves 2004, 2006, 2009 and 2011:

$$Nutrient_{n,head,j,t} = \beta_1 \, EDR_j + \sum_{v=1}^{V} \beta_v X_{v,h,j,t} + \sum_{t=2004}^{2011} \theta_t + \varepsilon_{h,j,t} \qquad (1)$$

Where $Nutrient_{n,h,j,t}$ denotes Nutrient intake *n* per head of households *head* located in province *j.* In particular, in my analysis I will examine the following caloric intake measured by kcal and macro-nutrients: Carbohydrates, Fat and Protein intake, where measured in grams.

X_h is a vector *v* of *i* time-varying household socio-economic variables, such as the logarithm of household total income, household size, completed years of education of the head of the household, gender, age and marital status of the head of the household, ethnicity, whether household is located in an urban or rural environment, and whether they maintain food habits from their parents' home. As a proxy for exposure to famine, I use EDR_j, which captures the excess death rate in province *j* between years 1959 and 1961. Equation (1) also includes $\sum_{t=2004}^{2011} \theta_t$ a vector of time-period (wave) dummies.

While instructive, a potential concern with the tests shown in Equation (1) is that it only looks at macro-nutrient intake in absolute terms. To address this issue, I re-estimate Equation (1) to now look at the relative share of each macro-nutrient in the overall diet:

$$Nutrient\_share_{n,head,j,t} = \beta_1 \, EDR_j + \sum_{v=1}^{V} \beta_v \, X_{v,h,j,t} + \sum_{t=2004}^{2011} \theta_t + \varepsilon_{h,j,t} \tag{2}$$

Where the share of *Nutrient n={carbohydrates, fat, protein}* per head of households *head* located in province *j* is defined as:

$$Nutrient\_share_{n,head,j} = \frac{Nutrient_n}{\sum_{n=1}^{3} Nutrient_n} \tag{3}$$

3.4.1.2 Determinants of macro-nutrient consumption: birth-cohort analysis

A potential concern with the above analysis is that it only captures an average effect and that it does not explicitly take into account the age of the household head at the time of the famine and also at the time of measurement of macro-nutrient intake. In this subsection I address this issue.

Based on existing population health literature (Almond et al., 2010; Cheng & Smyth, 2021; Herman et al., 2014), I analyse whether the exposure to famine during the prenatal and early periods of childhood will exert larger effects than at other subsequent periods. Given that Chapter One showed that the effect of famine on DDS was more pronounced on early childhood cohorts, I restrict the sample of birth cohorts to five years before and after the famine period. As a result, the control group is made up of those individuals who were born after the famine (i.e., from 1963 to 1969), and those born between 1954 and 1958[5].

I first quantify the lasting effects of the famine on macro-nutrient intake by estimating the following equation:

$$Nutrient\_share_{n,head,j,t} = \beta_1 \, EDR_j + \sum_{k=1954}^{1962} \gamma_k (EDR_j \times BirthYear_{head,k}) +$$

$$\delta_k + \sum_{v=1}^{V} \beta_v \, X_{v,h,j,t} + \sum_{t=2004}^{2011} \theta_t + \varepsilon_{h,j,t} \tag{4}$$

[5] My empirical results remain qualitatively and quantitatively similar if we include more birth cohorts prior to 1954 or after 1967.

where $Nutrient\_share_{n,head,j}$ is defined as above. δ_k are the cohort dummy variable, and EDR_j are the excess death rate of region *j* as defined above. $BirthYear_{head,k}$ is a dummy variable indicating whether *head of household (head) was born in the year k*. Note that *k=1962* refers to a birth cohort in gestation in 1961 and born in 1962. I include this birth cohort because, as discussed in the previous section, the exposure to famine during the foetal period may exert significant long-term effects. I expect that the magnitude of these estimated coefficients varies with birth cohorts. More specifically, I expect a larger impact of the famine on the relatively young birth cohorts, in particular those who were in gestation and early childhood during the famine.

3.4.2 The effect of famine on dietary diversity: importance of macro-nutrient consumption

In the final step, I examine the effect of early-life exposure to famine, as proxied by EDR, on dietary diversity later in life, while controlling for the relative share of each macro-nutrient in the overall diet:

$$D_{h,j,t} = \beta_1 EDR_j + \beta_2 Nutrient\_share_{h,j,t} + \sum_{i=1}^{I} \beta_i X_{i,h,j,t} + \sum_{t=2004}^{2011} \theta_t + \varepsilon_{h,j,t} \quad (5)$$

Where $D_{h,j,t}$ denotes dietary diversity score of household *h* located in province *j.* Other variables are defined as above.

3.5 Results

The existing literature tries to model dietary patterns using various socio-economic variables. Education, income, ethnicity, urban residence, marital status, nutritional and dietary knowledge have found to be positively associated with dietary diversify score, while smoking and alcohol consumption have been negatively associated to DDS (Hou et al., 2021; Wang et al., 2021; Zhang et al., 2020; Zhang & Zhao, 2021; Zhang et al., 2017; Zhao et al., 2020; Zhao et al., 2017; Zhong et al., 2018). On the other hand, evidence on the impact of adverse events such as famine has not been explored so far. Evidence presented in *Section Two* suggests that exposure to extreme life changing events, such as famine has long-lasting effects on food choice, and in particular on dietary diversity score. In this *Section* I examine potential channels through which early-life exposure to famine affects nutrient intake and food choices later in life.

3.5.1 Early life famine and food intake

Analysis presented in the previous subsections suggests that early life famine had a differential effect on dietary diversity score later in life, based on individual's income, educational attainment and gender. In particular, it highlights that among famine survivors, high-income, high-educated male individuals tend to have higher dietary diversity later in their lives. This result is important, since it sheds more light on the established relation between income and education on one side, and dietary diversity on the other, as it highlights that this relation is driven by those individuals who had lower previous exposure to extreme events such as famine. In this *Section*, I expand this analysis to look at the effects of early life famine on food and macro-nutrient intake later in life.

In Table 2.2 I show the results of estimating Equation (1) on the pooled sample that includes four waves {2004, 2006, 2009 and 2011}. Columns (1)-(4) show the results of estimating Equation (1) without the quadratic term *Age*^{2} *(Head),* while columns (5)-(8) show the results of estimating the full Equation (1). The dependent variable is average household head's *Kcal* intake (columns 1 and 4), *Carbohydrate* intake measured in grams (column 2 and 5), *Fat* intake measured in grams (columns 3 and 7) and *Protein* intake measured in grams (columns 4 and 8).

As before, all specifications contain a vector of covariates that includes: natural logarithm of household total income, household size, (household head's) completed years of education, age, gender and marital status of the head of the household, (household head's) nationality, whether household is located in an urban or rural environment, and whether they maintain food habits from their previous home. All specifications include wave dummies to control for the general macro-economic conditions that are affecting the sample population at every time period, as well as robust standard errors that are adjusted for heteroscedasticity.

The main independent variable of interest *EDR* captures the excess death rate *EDR[%]* in the period 1959-1961 in province *j*, where the household is located. As we can see from *Table 2.2*, the estimated coefficients on excess death rate *EDR* are negative and significant at 1% level in all specifications, suggesting that individuals who had a high prior exposure to famine, or those who live in areas more severely affected by famine tend to consume less calories, carbohydrates, fat and protein in their diet in absolute terms.[6]

A close inspection of the estimated coefficients on the covariates suggests that higher income households (individuals) tend to consume more calories, fat and protein in their diet (the estimated coefficients are positive and significant at 1% level), and fewer carbohydrates (the estimate coefficient in Column 2 and 6 is negative, although not significant at conventional levels). Estimated coefficient on *Age (Head)* is negative and significant at 1% level for Kcal, Carb and Protein intake, and negative and significant at 10% level for Fat intake (columns 1-4), suggesting that older individuals tend to reduce their food intake across the board. Interestingly, when I include the quadratic term *Age^2 (Head)* in Columns (5)-(8), estimated coefficient on *Age (Head)* becomes positive (although not significant) for *Kcal*, *Carbs* and *Protein*, and remains negative for *Fat* intake. At the same time, estimated coefficient on the quadratic term *Age^2 (Head)* enters with a negative sign in all four columns, being statistically significant at 1% level for *Kcal,* 10% level for *Carbs*, 5% level for *Protein* and not significant for *Fat*. This result suggests that the negative relation between individual's age and their macro-nutrient intake is concave and driven by the very oldest part of the population.

[6] Results remain quantitatively unchanged when I cluster standard errors by province (see table A.2.1)

Table 2. 2: Effects of famine on absolute intake of calories and nutrients; Pooled regressions (2004 -2011)
Dependent variables: Kcal presents presents absolute intake of calories of household head per day; Carbs presents absolute intake of carbohydrates of household head per day (g/capita/day); Fat presents absolute intake of fat of household head per day (g/capita/day); Protein intake presents absolute intake of protein of household head per day (g/capita/day)
Explanatory variables: log(HH total income) presents logarithmic value of Household total income; Education, EDR [%], Age, Gender, HH Size, Han nationality, 'Old home' food habits, Marital Status Urban is defined as in Table 1.2
*Robust standard errors are adjusted for heteroscedasticity; Significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; All specifications include wave dummy variables.*

	(1) Kcal	(2) Carbs	(3) Fat	(4) Protein	(5) Kcal	(6) Carbs	(7) Fat	(8) Protein
log(HH total income)	64.101***	-6.011	5.442***	3.767***	60.708***	-6.386	5.434***	3.665***
	[2.729]	[-1.486]	[4.134]	[4.343]	[2.584]	[-1.577]	[4.105]	[4.240]
Education	-5.547	-2.461***	0.801***	-0.124	-5.787	-2.487***	0.801***	-0.131
	[-1.305]	[-3.359]	[3.635]	[-0.810]	[-1.363]	[-3.400]	[3.628]	[-0.858]
EDR [%]	-529.473***	-75.268***	-15.273***	-20.958***	-513.949***	-73.553***	-15.236***	-20.494***
	[-9.829]	[-8.500]	[-5.648]	[-10.804]	[-9.490]	[-8.229]	[-5.620]	[-10.488]
Age	-1,282.404***	-230.470***	-14.695*	-45.858***	1,933.703	124.864	-6.964	50.386
	[-7.959]	[-8.469]	[-1.690]	[-8.109]	[1.565]	[0.581]	[-0.101]	[1.003]
Age^2					-2,760.348***	-304.979*	-6.636	-82.605**
					[-2.627]	[-1.675]	[-0.113]	[-1.980]
Gender	418.888***	71.731***	0.565	11.923***	422.994***	72.185***	0.575	12.046***
	[8.695]	[8.582]	[0.218]	[7.037]	[8.809]	[8.649]	[0.221]	[7.108]
HH size	12.619	7.918***	-2.144***	0.691*	7.220	7.322***	-2.157***	0.530
	[1.297]	[4.843]	[-4.222]	[1.917]	[0.730]	[4.395]	[-4.151]	[1.395]
Han nationality	-111.128***	-24.895***	1.800	-2.663*	-99.712**	-23.634***	1.828	-2.322*
	[-2.804]	[-3.925]	[0.894]	[-1.947]	[-2.507]	[-3.707]	[0.900]	[-1.691]
'Old home' food habits	80.469	11.382	4.812	1.230	80.065	11.337	4.811	1.218
	[1.224]	[1.058]	[1.532]	[0.497]	[1.220]	[1.055]	[1.532]	[0.491]
Marital Status	-109.751*	-39.421***	4.943	-0.227	-116.929*	-40.214***	4.926	-0.442
	[-1.711]	[-2.950]	[1.504]	[-0.097]	[-1.821]	[-3.015]	[1.500]	[-0.190]
Urban	-138.453***	-26.240***	-3.819	-2.825	-139.317***	-26.335***	-3.821	-2.851
	[-2.853]	[-3.427]	[-1.569]	[-1.302]	[-2.877]	[-3.446]	[-1.570]	[-1.313]
Wave dummies	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3,283	3,283	3,283	3,283	3,283	3,283	3,283	3,283
R-squared	0.118	0.123	0.063	0.107	0.120	0.124	0.063	0.108
AIC	52,352	40,646	33,146	30,621	52,348	40,645	33,148	30,619
BIC	52,438	40,731	33,231	30,707	52,440	40,737	33,239	30,711

Male individuals tend to have higher *Kcal, Carb* and *Protein* intake, as indicated by the positive and statistically significant at 1% level coefficients. Their Fat intake is also higher, although this coefficient is not precisely estimated (columns 3 and 7). Larger households, as indicate by coefficients on *HH size*, tend to intake more *Carbs* and *Protein* (estimated coefficients are statistically significant at 1% and 10% level respectively), and less *Fat* (estimated coefficient is statistically significant at 1% level). Married heads of household tend to consume less *Kcal* and fewer *Carbs*, as indicated by negative and statistically significant coefficient estimates (10% and 1% level, respectively). Finally, individuals located in *Urban* areas also tend to consume less *Kcal* and fewer *Carbs*, as indicated by negative and statistically significant coefficient estimates (1% level).

As we can see in Table 2.2, adding Age^2 to the model improves its relative goodness-of-fit when measured by AIC, where dependent variable is *kCal*, *Carbs* and *Protein*, but not *Fat*. When relative goodness-of-fit is measured by BIC, adding Age^2 does not improve the model in any specification.

Results presented above indicate that prior exposure to famine reduces macro-nutrient intake in absolute sense later in life. In the next test, I explore whether early life exposure to famine changes to composition of individual's diet, that is, whether it affects the *relative share* of *Carb, Fat* and *Protein* intake in the overall individual's food intake. In particular, I re-estimate Equation (1), where now my main dependent variables are the relative share of each macro-nutrient in the overall diet:

$$Nutrient\_share_{n,h,j} = \frac{Nutrient_n}{\sum_{n=1}^{3} Nutrient_n}$$

Where $MacroNutritient\_Intake_{i,h}^{j}$ represents the average household head' *h*'s intake of macro-nutrient *n={Carbs, Fat, Protein}* for each household *h*. Results are reported in *Table 2.3*.

The estimated coefficient on the main independent variable of interest *EDR* is positive and significant at 1% level for *Carb_share*, ranging from 0.020 in Column 1 to 0.021 in Column 4, suggesting that individuals who had a high prior exposure to famine tend to consume a higher share of carbohydrates in their overall diet. The estimated coefficient on the main independent variable of interest *EDR* is negative and significant at 1% level for *Protein_share*, ranging from -0.011 in Column 2 to -0.011 in Column 5, suggesting that individuals who had a high prior exposure to famine tend to consume a lower share of protein in their overall diet. Looking at the estimated coefficients on *EDR* for *Fat share* as the main dependent variable (columns 3 and 6), we see that they are also negative and significant at 10% level, suggesting that early life famine leads to lower *Fat share* in the overall diet later in life. Results in the table 2.2 indicate that exposure to famine is positively correlated with higher share of carbohydrates in the overall diet. This can potentially be driven by higher cereal consumption. Also, the results from Table 2.3 suggest that there is a negative correlation between exposure to famine and fat and protein intake. This can potentially be driven by lower consumption of animal sourced foods.

Table 2. 3: Effects of famine on relative share of macronutrients in total diet; Pooled regressions (2004-2011)
Dependent variables: Carb share presents share of carbohydrates in total diet of household head (%); protein share presents share of protein in total diet of household head (%); fat share presents share of fat in total diet of household head (%)
Explanatory variables as defined in Table 2.2.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; All specifications include wave dummy variables.*

	(1)	(2)	(3)	(4)	(5)	(6)
	Carb share	Protein share	Fat share	Carb share	Protein share	Fat share
log(HH total income)	-0.015***	0.007***	0.008***	-0.015***	0.007***	0.009***
	[-5.111]	[6.105]	[3.253]	[-5.155]	[6.019]	[3.349]
Education	-0.002***	0.000*	0.002***	-0.002***	0.000*	0.002***
	[-4.711]	[1.903]	[4.709]	[-4.734]	[1.864]	[4.763]
EDR [%]	0.020***	-0.011***	-0.009*	0.021***	-0.011***	-0.010*
	[3.218]	[-4.592]	[-1.664]	[3.327]	[-4.387]	[-1.886]
Age	-0.046**	-0.006	0.052***	0.118	0.089	-0.207
	[-2.422]	[-0.866]	[3.076]	[0.753]	[1.471]	[-1.472]
Age^2				-0.140	-0.082	0.222*
				[-1.037]	[-1.570]	[1.826]
Gender	0.027***	0.000	-0.027***	0.027***	0.000	-0.027***
	[4.006]	[0.019]	[-4.523]	[4.038]	[0.065]	[-4.583]
HH size	0.007***	-0.001	-0.006***	0.006***	-0.001*	-0.006***
	[6.127]	[-1.515]	[-6.333]	[5.799]	[-1.863]	[-5.835]
Han nationality	-0.009*	0.001	0.008*	-0.008*	0.001	0.007*
	[-1.858]	[0.579]	[1.911]	[-1.724]	[0.774]	[1.671]
'Old home' food habits	-0.005	-0.001	0.007	-0.005	-0.001	0.007
	[-0.758]	[-0.434]	[1.118]	[-0.762]	[-0.437]	[1.124]
Marital Status	-0.024***	0.009***	0.015**	-0.024***	0.009***	0.015**
	[-2.817]	[3.161]	[2.040]	[-2.878]	[3.084]	[2.151]
Urban	-0.011*	0.004*	0.007	-0.011*	0.004*	0.007
	[-1.890]	[1.776]	[1.416]	[-1.900]	[1.763]	[1.434]
Wave dummies	Y	Y	Y	Y	Y	Y
Observations	3,283	3,283	3,283	3,283	3,283	3,283
R-squared	0.091	0.071	0.068	0.091	0.071	0.069
AIC	-7,078	-13,396	-8,109	-7,077	-13,397	-8,111
BIC	-6,992	-13,311	-8,024	-6,985	-13,306	-8,019

Similarly to Table 2.2 when I explore impact of EDR on share of macronutrients in diet, AIC test suggests that adding Age^2 improves relative goodness-of-fit in case of Protein share and Fat share, but not in case of Carb share. BIC on the other hand suggests that adding Age^2 does not improve relative goodness-of-fit of any of the 3 specifications.

3.5.2 Dietary diversity score, macro-nutrient share and early life famine

Evidence presented above suggest that early life famine has long lasting effects on carbohydrate, fat and protein share in an individual's diet. In the next set of tests, I relate this finding to dietary diversity. In particular, I estimate Equation (5), which models dietary diversity score as a function of early life exposure to famine and relative share of macro-nutrients (carbohydrates, fat and protein), using the pooled sample containing four waves {2004, 2006, 2009 and 2011}. The vector of covariates is the same as before. All specifications include wave dummies to control for the general macro-economic conditions that are affecting the sample population at every time period, as well as robust standard errors that are adjusted for heteroscedasticity. Results are shown in *Table 2.4*.

Estimated coefficient on *Carb share* is negative and significant at 1% level in Columns (1) and (2), suggesting that individuals with lower relative share of carbohydrates tend to have higher dietary diversity scores. Estimated coefficient on *EDR* reported in column (2) is also negative and significant at 1% level, suggesting that controlling for the relative share of carbohydrates, lower prior exposure to famine leads to higher dietary diversity scores.

Table 2. 4: Effects of share of macronutrients in diet on Dietary Diversity Score, Pooled regressions (2004-2011)

Dependent variable: Dietary Diversity Score is a constructed variable based on 20 food groups – theoretical range of this value is from 0.1 to 1

Explanatory variables as defined in Table 2.2.; Carb share, Protein share and Fat share presents share of each macronutrient in total diet of household head (%)

*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; All specifications include wave dummy variables.*

	(1)	(2)	(3)	(4)	(5)	(6)
	Dietary Diversity Score					
log(HH total income)	0.056***	0.054***	0.054***	0.052***	0.060***	0.058***
	[14.637]	[14.229]	[14.229]	[13.866]	[15.262]	[14.825]
Education	0.001**	0.001	0.002***	0.001**	0.002***	0.001*
	[2.482]	[1.358]	[3.492]	[2.408]	[3.046]	[1.851]
Carb share	-0.446***	-0.435***				
	[-19.285]	[-19.097]				
EDR [%]		-0.072***		-0.067***		-0.078***
		[-10.509]		[-9.946]		[-10.952]
Age	-0.007	-0.012	0.021	0.015	-0.004	-0.010
	[-0.301]	[-0.545]	[0.971]	[0.717]	[-0.193]	[-0.455]
Gender	0.018**	0.012*	0.006	0.000	0.017**	0.010
	[2.567]	[1.670]	[0.867]	[0.046]	[2.248]	[1.326]
HH size	-0.006***	-0.003**	-0.008***	-0.006***	-0.007***	-0.004***
	[-4.324]	[-2.537]	[-6.080]	[-4.287]	[-4.861]	[-2.992]
Han nationality	0.010**	-0.021***	0.011**	-0.018***	0.014***	-0.020***
	[2.443]	[-4.201]	[2.535]	[-3.739]	[3.243]	[-3.918]
'Old home' food habits	0.001	0.006	0.006	0.010	0.001	0.006
	[0.130]	[0.650]	[0.577]	[1.056]	[0.088]	[0.623]
Marital Status	0.032***	0.026***	0.030***	0.025***	0.037***	0.032***
	[4.031]	[3.365]	[3.711]	[3.120]	[4.400]	[3.737]
Urban	0.024***	0.027***	0.024***	0.027***	0.026***	0.030***
	[3.104]	[3.575]	[3.141]	[3.606]	[3.309]	[3.799]
Protein share			1.262***	1.222***		
			[17.729]	[17.508]		
Fat share					0.358***	0.351***
					[12.827]	[12.780]
Wave dummies	Y	Y	Y	Y	Y	Y
Observations	3,283	3,283	3,283	3,283	3,283	3,283
R-squared	0.266	0.286	0.284	0.301	0.214	0.237
AIC	-5,793	-5,880	-5,870	-5,947	-5,567	-5,662
BIC	-5,708	-5,789	-5,785	-5,856	-5,481	-5,570

Estimated coefficient on *Protein share* is positive and significant at 1% level in Columns (3) and (4), suggesting that individuals with higher relative share of protein tend to have higher dietary diversity scores. Estimated coefficient on *EDR* reported in column (4) is remains negative and significant at 1% level, suggesting that controlling for the relative share of protein, lower prior exposure to famine leads to higher dietary diversity scores.

Estimated coefficient on *Fat share* is positive and significant at 1% level in Columns (5) and (6), suggesting that individuals with higher relative share of fat in their diet tend to have higher dietary diversity scores. Estimated coefficient on *EDR* reported in column (6) is remains negative and significant at 1% level, suggesting that controlling for the relative share of fat, lower prior exposure to famine leads to higher dietary diversity scores.

As we can see in table 2.4, AIC and BIC reveal consistent findings. Adding ERD[%] to the specification, improves relative goodness-of-fit of the models which measure impact of Carb share and Protein share on Dietary Diversity Score, while the same cannot be claimed for the impact of Fat share on Dietary Diversity Score.

3.5.3 Macro-nutrient share and early life famine: birth-year cohort analysis

I next quantify the lasting effects of the famine on macro-nutrient intake of the survivors by birth-year cohort by re-estimating Equation (4), where dependent variable is now nutrient $Nutrient_{head,j}$.

Note that k=1962 refers to a birth cohort in gestation in 1961 and born in 1962. We include this birth cohort because, as discussed in the previous section, the exposure to famine during the foetal period may exert significant effects later in life.

The coefficient of the interaction between the excess death rate and birth cohort dummy variables measures the effect of the famine on macronutrient intake. I expect that the magnitude of these estimated coefficients varies with birth cohorts. More specifically, based on the earlier analysis, I expect a larger impact of the famine on the relatively young birth cohorts, especially those who were in early childhood during the famine (2-5 YO). More specifically, based on the results of *Section Two* (*Table A.1.11*) and on the previous results in this *Section* I expect coefficient in year cohort 1957 to be significant and positive for carbohydrates and significant and negative for fat and protein. The total sample includes those individuals born between 1954 and 1962 as the treatment group, and individuals born between 1963 and 1967, which are treated as the control group.

Table 2. 5: Effects of famine on absolute intake of calories and nutrients of birth cohorts 1954-1962; Pooled regressions (2004-2011)

Dependent variables: as described in Table 2.1

*Explanatory variables: Birthyear1954xEDR presents interaction term between birth year cohort 1954 and excess death rate; interaction terms of birth cohorts 1955-1962 follow the same pattern; other explanatory variables as described in Table 2.2 Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; All specifications include wave dummy variables and birth-cohort dummies which are supressed for brevity.*

	(1) Kcal	(2) Carbs	(3) Fat	(4) Protein
log(HH total income)	78.400*	-9.870	7.601***	3.879**
	[1.746]	[-1.386]	[3.080]	[2.206]
Education	4.651	-0.597	0.631	0.216
	[0.619]	[-0.443]	[1.572]	[0.827]
Gender	587.267***	85.527***	10.406*	21.289***
	[5.750]	[4.873]	[1.884]	[5.971]
HH size	28.784	18.850***	-3.780***	1.131
	[1.406]	[5.185]	[-3.334]	[1.448]
Han nationality	-20.707	-15.873	7.866**	1.618
	[-0.283]	[-1.270]	[2.159]	[0.602]
'Old home' food habits	78.899	12.933	-1.167	-2.488
	[0.631]	[0.630]	[-0.189]	[-0.499]
Marital Status	-34.576	-28.779	11.109*	1.473
	[-0.284]	[-1.234]	[1.918]	[0.318]
Urban	-74.327	-24.946*	-1.202	0.434
	[-0.827]	[-1.852]	[-0.247]	[0.096]
EDR [%]	-106.430	-25.798	-2.887	-8.522
	[-0.422]	[-0.695]	[-0.230]	[-1.072]
Birthyear1954 x EDR [%]	-284.630	-23.814	-3.939	0.829
	[-0.774]	[-0.396]	[-0.248]	[0.063]
Birthyear1955 x EDR [%]	-508.017	-77.875	-15.686	-10.102
	[-1.639]	[-1.562]	[-0.998]	[-0.968]
Birthyear1956 x EDR [%]	-648.161**	-62.856	-18.913	-18.554*
	[-2.117]	[-1.289]	[-1.200]	[-1.842]
Birthyear1957 x EDR [%]	-433.178	-43.455	-13.460	-11.456
	[-1.365]	[-0.752]	[-0.744]	[-1.089]
Birthyear1958 x EDR [%]	-482.788	-54.049	-9.742	-3.789
	[-1.030]	[-0.719]	[-0.446]	[-0.177]
Birthyear1959 x EDR [%]	-572.515	-113.745*	8.818	-23.267
	[-1.384]	[-1.704]	[0.464]	[-1.076]
Birthyear1960 x EDR [%]	-1,035.207*	-261.276***	1.585	-33.709*
	[-1.667]	[-3.419]	[0.046]	[-1.707]
Birthyear1961 x EDR [%]	-487.375	-57.998	-10.971	-13.370
	[-1.070]	[-0.638]	[-0.473]	[-1.009]
Birthyear1962 x EDR [%]	-1,152.937*	-343.577***	19.967	-58.745**
	[-1.839]	[-3.384]	[0.625]	[-2.404]
Wave dummies	Y	Y	Y	Y
Observations	1,064	1,064	1,064	1,064
R-squared	0.109	0.144	0.075	0.093
AIC	17,015	13,198	10,798	10,067
BIC	17,169	13,352	10,952	10,222

Columns (1)–(4) of *Table 2.5* present the regression results of estimating the long-term effect of famine on macro-nutrient intake later in life, whereby each column presents the results of pooled cross-sectional regressions for the pooled sample comprised of the four waves {2004, 2006, 2009 and 2011}. Note that the dummy variables for birth cohorts from 1954 to 1962 are controlled in the regressions in *Table 2.5*, but their coefficients are not reported due to space limitations. All specifications include wave dummies, and standard errors are heteroscedasticity-adjusted. I mainly report the coefficients of interaction terms between excess death rate *EDR* [%] and birth cohorts, which, as discussed above, measure the estimated effects of the famine on macro-nutrient intake. The estimated coefficient on EDR is negative in all four columns, although not precisely estimated.

Most famine-affected birth cohorts have lower *Kcal* intake than the counterfactual case that the famine had not occurred, which is reflected in the negative coefficients of the interaction terms *EDRxBirthYear*. This implies that the famine generally caused adverse effect on the *Kcal* intake of the survivors. Among the cohorts with adverse effects of famine, the birth cohorts of 1956, 1960 and 1962 have more severe and also statistically significant effects. As evidenced in Column 1, the estimated coefficient on the interaction term is negative and statistically significant at 5% (1%) level respectively. These results suggest that exposure to famine during early years of childhood (i.e. for children who were around 2 years of age when the famine started), and those who were in gestation age during the famine (those born in 1960), gives rise to more devastating long-term effects in later life, which is consistent with general findings in population health literature (e.g., Barker, 1989, 1992; Heymann et al., 2005).

In Column 2, the estimated coefficient on the interaction terms *EDRxBirthYear* is negative for most birth years, and statistically significant for those born in 1959 (10% level), 1960 (1% level) and 1962 (1% level), suggesting that those birth cohorts had a more severe effect of famine on their carbohydrate intake in absolute terms. Similar results are also obtained in Column 4, where the dependent variable is *Protein* intake in grams: the estimated coefficient on the interaction terms *EDRxBirthYear* is negative for most birth years, and statistically significant for those born in 1956 (10% level), 1960 (1% level) and 1962 (5% level), suggesting that those birth cohorts had a more severe effect of famine on their protein intake in absolute terms. Interestingly, the effect of famine on *Fat* intake is a bit more nuanced, as it can be seen from Column 3. Estimated coefficients on the interaction terms *EDRxBirthYear* are negative for most birth years, with the exception of those born in 1960 and in 1962, whose estimated coefficient on the interaction term is positive, although not statistically significant.

In *Table 2.6*, we repeat the analysis by re-estimating Equation (4), where the dependent variable is now the relative *share* of each macro-nutrient in the individual's overall diet. Columns (1)–(3) of *Table 2.6* present the regression results of estimating the long-term effect of famine on macro-nutrient *share* later in life, whereby each column presents the results of pooled cross-sectional regressions for the pooled sample comprised of the four waves {2004, 2006, 2009 and 2011}. Note that the dummy variables for birth cohorts from 1954 to 1962 are controlled in the regressions in *Table 2.6* but their coefficients are not reported due to space limitations. All specifications include wave dummies, and standard errors are heteroscedasticity-adjusted. We mainly report the coefficients of interaction terms between excess death rate *EDR* and birth cohorts, which, as discussed above, measure the estimated effects of the famine on macro-nutrient intake. The estimated coefficient on *EDR* is negative for *Protein* and *Fat* share, and positive for *Carb* share, although not precisely estimated.

Table 2. 6: Effects of famine on relative share of macronutrients in total diet of birth cohorts 1954-1962; Pooled regressions (2004-2011)

Dependent variables: as described in Table 2.1

Explanatory variables: as described in Table 2.2

*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; All specifications include wave dummy variables and birth-cohort dummies which are supressed for brevity.*

	(1) Share of carbs	(2) Share of protein	(3) Share of fat
log(HH total income)	-0.017***	0.008***	0.009**
	[-3.052]	[3.424]	[1.970]
Completed Years Education (Head)	-0.001	0.000	0.001
	[-1.639]	[0.979]	[1.488]
Gender (Head)	0.006	0.012**	-0.018
	[0.433]	[2.303]	[-1.415]
HH size	0.014***	-0.002***	-0.012***
	[5.764]	[-2.854]	[-5.269]
Han nationality (Head)	-0.018*	0.003	0.015*
	[-1.899]	[0.793]	[1.857]
Old home food habits	0.010	-0.006	-0.005
	[0.738]	[-1.097]	[-0.394]
Marital Status (Head)	-0.036**	0.007	0.029**
	[-2.406]	[1.114]	[2.561]
Urban	-0.017	0.007*	0.009
	[-1.614]	[1.740]	[1.074]
EDR [%]	0.001	-0.005	0.004
	[0.035]	[-0.458]	[0.163]
Birthyear1954 x EDR [%]	-0.008	0.005	0.003
	[-0.212]	[0.281]	[0.093]
Birthyear1955 x EDR [%]	0.015	0.001	-0.016
	[0.429]	[0.074]	[-0.536]
Birthyear1956 x EDR [%]	0.028	-0.016	-0.012
	[0.777]	[-1.216]	[-0.402]
Birthyear1957 x EDR [%]	0.005	-0.003	-0.002
	[0.126]	[-0.154]	[-0.066]
Birthyear1958 x EDR [%]	0.026	0.003	-0.029
	[0.426]	[0.123]	[-0.607]
Birthyear1959 x EDR [%]	-0.032	-0.016	0.047
	[-0.727]	[-0.516]	[1.129]
Birthyear1960 x EDR [%]	-0.080	0.016	0.064
	[-1.395]	[1.002]	[1.258]
Birthyear1961 x EDR [%]	0.001	-0.010	0.009
	[0.015]	[-0.730]	[0.166]
Birthyear1962 x EDR [%]	-0.106	0.003	0.102**
	[-1.619]	[0.116]	[1.989]
Wave dummies	Y	Y	Y
Observations	1,064	1,064	1,064
R-squared	0.126	0.094	0.099
AIC	-2,234	-4,300	-2,564
BIC	-2,080	-4,146	-2,410

In Column (1), the estimated coefficient on the interaction terms *EDRxBirthYear* is negative for most birth years, while being positive for those born in 1955, 1956, 1957, 1958 and 1961, suggesting that those birth cohorts are driving the results reported in Column (1) and (2) of *Table 2.3*, that suggests that early life exposure to famine is associate with higher relative share of carbohydrates later in life. It is important to note, however, that most of these coefficients have large standard errors, and are not statistically different from zero.

Similar results are also obtained in Column (2), where the dependent variable is *Protein share*: the estimated coefficient on the interaction terms *EDR x BirthYear* is positive for most birth years, while being negative for those born in 1956, 1957, 1959 and 1961, suggesting that those birth cohorts are driving the results reported in Column 3 and 4 of *Table 2.3*, that suggests that early life exposure to famine is associated with lower relative share of protein later in life. The effect of famine on *Fat share* is also nuanced, as it can be seen from Column 3. Estimated coefficients on the interaction terms *EDR x BirthYear* are positive for most birth years, with the exception of those born in 1955, 1956, 1957 and 1958, whose estimated coefficient on the interaction term is negative, although not statistically significant. This suggests that early life exposure to famine is associated with lower relative share of fat later in life. Interestingly, the estimated coefficient on the interaction term is positive and significant for those born in 1962, suggesting that for this birth cohort, exposure to famine during gestation age had a positive effect on the relative *Fat share* in their overall diet later in life.

Taken together, these results demonstrate large and devastating effects of the great famine on the surviving population which occurred 50 years before the waves used in the analysis.

4 Long-term effects of famine on lifestyle choices

Previous two sections explored the effects of exposure of famine on qualitative and quantitative characteristics of diet. Qualitative characteristics were proxied by Dietary Diversity Score (DDS) and share of individual macronutrients in total diet, where macronutrients are fat, carbohydrates and protein. Quantitative characteristics of diet were proxied by daily calories intake, as well as daily intake of fat, carbohydrates, and protein, measured in grams. My results show that both qualitative and quantitative characteristics of diets are negatively correlated with exposure to famine.

In this section, I explore relationship between exposure to food shortage and famine on lifestyle choices, which are directly or indirectly connected to food consumption. The lifestyle choices are proxied by household expenditures on different items, such as total household food expenditures and expenditures on food eaten away from home. Additionally, I analyse effect of famine on family planning decisions, as those decisions are indirectly associated with nutritive intake of children of those who make the family planning decisions. By applying this approach, not only I aim to confirm my findings related to impacts of famine on nutrient intake, from previous two sections, but I also aim to add another dimension to area which explores long-term effects of food shortage, and its most severe form – famine starvation.

4.1 Evidence from the existing literature

To analyse the literature which examined the impact of extreme events, such as famine, earlier in life on lifestyle choices later in life, it is necessary to define lifestyle choices. The current literature does not provide a unique definition. Approaches to the concept of lifestyle could be Weberian, sub-cultural, psychological, market research and psychographics, leisure/tourism styles, spatial, socialist lifestyles, consumer culture, gender, as well as a miscellaneous group, and the application of those approaches depend on the discipline which is analysing lifestyles (Veal, 1993). In the Weberian approach, the emphasis is not only on the societal divisions based on class, but also on the status, which is honour based (Veal, 1993). While Weberian approach analyses lifestyles from societal - group perspective, psychological and consumer culture approach analyses lifestyles from an individual perspective. Psychological approach, like the Weberian one, analyses guiding values and principles and frames individuals as a whole person, while consumer culture approach deals with notion of "conflict" between real and illusional freedom of choice of consumers in capitalist, post-modern societies (Veal, 1993). Another very important dimension of lifestyle debate is structured around coherence. Namely, "Diderot Effect" posits that there is a force which leads individuals to maintain a cultural consistency when it comes to obtaining consumer goods, so that the goods are complementary (Veal, 1993). This might imply that consumers are obtaining goods and services in such a way that it fits a certain lifestyle, perceived either by themselves, or by others.

Similarly, other researchers analysed the concept of lifestyle from pluralistic perspective. It has been argued that the lifestyle is mere self-expression of an individual and it can include eating and drinking, one's preferred means of transportation, political engagement, and others (Jensen, 2007). Furthermore, in analysing sustainable development issues such as consumerism and greenhouse gases, the author argues that lifestyle concept can be used on four levels: global, national, sub-cultural and the individual level (Jensen, 2007). Also, analysis of American culture which argues that consumption of goods and services is encouraged from the early childhood, also perceives the consumption as a reflection of structural position, preferences and identity (Keister et al., 2016). Based on the expenditures, authors then cluster population sample in different groups. Another piece of research which also deals with population segmentation based on lifestyles, argues that in addition to the traditional basis for segmentation, which are activities, interests and opinions, it is necessary to add attributes which reflect lasting personal characteristics, and those are values, life visions, aesthetic style and media preferences (Vyncke, 2002).

The current literature which explores the effects of certain lifestyle choices on health refers to those as lifestyle-risk factors, when they contribute to certain conditions. Those risk factors are tobacco, alcohol use or sedentary lifestyle (Carballo-Fazanes et al., 2020; Dieteren & Bonfrer, 2021; McPhee et al., 2016; Ruiz, 2021). On another hand, healthy eating, physical activity and are some of the lifestyle choices which are associated to a positive impact on health (Chadwick, 2013).

Finally, diversity of journals where the papers in this review have been published reflects the crosscutting nature of "lifestyle". Those are: Leisure Studies, Environmental Sciences, BMC Public Health, Biogerontology, International Journal of Environmental Research and Public Health, Sociological Science, European Journal of Communication, Health Education Quarterly, and Bioethics.

As nutrition transition theory postulates, when low- and medium-income countries make economic progress, notable shifts in diet and activity patterns are being observed (Popkin, 2006). Among other things, these changes lead to increased consumption of energy dense foods and decreased physical activity level, which all result in energy imbalance and rising obesity. Lifestyle choices that will be analysed in this dissertation are those connected to energy intake. Food away from home (FAFH) is known for its energy-dense characteristics and that is one of the reasons why it has been extensively studied. Some authors focus on nutritional characteristics of FAFH (Zang et al. 2018; Guenther et al. 1981), while others compare dietary outcomes from different categories of away from home eating venues such as schools, canteens, conventional restaurants and hawker centres and food courts (McCracken and Brandt 1987; Naidoo et al. 2017). Consumption rate of fast food has been associated with low physical activity level and low intake of fruits and vegetables (Dunneram et al. 2013). Socioeconomic and demographic characteristics which were often examined in FAFH studies and associated to FAFH frequency are employment status, income, age, gender, household size, seasonality, part of the week, number of indoor restaurants in the neighbourhood where respondents live, time value (Nayga et al. 1992; Tian et al. 2016; McCracken and Brandt 1987). Most of FAFH literature examine determinants which are concurrent with food consumption. One of very few studies which examines the relationship between habits in the past and food consumption later in life looks at the eating behaviour established early in life and persistence of those habits in adulthood (Campbell and Crawford 2001). The authors suggest that family environment is an important factor for adult life dietary habits.

Apart from FAFH, lifestyle analysis also includes consumption of alcohol, and tobacco. (Drobes, 2002). For both of these products, genetic predispositions and environment factors as triggers for genetic predispositions have been widely examined (Drobes, 2002; Dick and Kendler, 2012). Studies which examine adverse childhood experiences (ACE) and alcohol initiation and consumption suggest that that there is a relationship between the two, thus identifying another dimension of adult lifestyle choices which has been influenced by adverse events in childhood (Dube et al. 2006). By ACE, the authors refer to childhood abuse and neglect, growing up with various forms of household dysfunction and alcohol use in adolescence and adulthood, but it does not include famine. Importance of applying wholistic lifestyle approach in studies which examine the effects of alcohol consumption on obesity has been raised by Traversy and Chaput (2015), who argue that these studies lack physical activity and sedentary behaviour information, which might result in inadequate analysis.

Analysis of lifestyle through consumer-culture lenses, might reveal the relationship between famine and composition of household food expenditures. To the best of my knowledge, no research examined this particular relationship. Having said that, research which did examine the relationship between DDS and household expenditures found there is a positive and linear relationship between per capita total household expenditures as well as per capita household food expenditures and DDS (Thorne-Lyman et al., 2010). The same research finds that comparing food expenditures of the lowest and the highest wealth quintile, households belonging to the highest quintile spend more on all food groups, yet this difference is relatively modest in case of rice, it is slightly bigger in case of vegetables, and is much bigger in case of fish, meat, legumes, fruits and eggs (Thorne-Lyman et al., 2010).

In the same vein, an extensive body of literature explores determinants of household food expenditures. However, majority of the literature focus on food expenditures in rural or urban poor areas, or agricultural households. The most prominent determinants are income, household size, educational attainment of the respondent, and marital status. Certain findings of this type of research are consistent, while others provide inconclusive evidence. Results show that income and household size are positively correlated with household food expenditure (Babalola & Isitor, 2014; Habib et al., 2018; Rubhara et al., 2020; Sekhampu, 2012; Sotsha et al., 2019; Zani et al., 2019). Additionally, being married is negatively associated to food expenditures (Habib et al., 2018; Sekhampu, 2012). Results of the reviewed literature, suggest that the relationship between education of the respondent and household food expenditure is both positive (Sekhampu, 2012) and negative (Habib et al., 2018; Sotsha et al., 2019; Zani et al., 2019).

As the reviewed literature suggests, there are very few attempts to examine childhood circumstances and their effect on adult lifestyle choices. Additionally, none of the reviewed articles examined the relationship between extreme events and lifestyle choices. Furthermore, several studies examined both food intake and physical activity, or food intake and tobacco and alcohol consumption, where one study pointed at shortcomings in research which do not take several lifestyle choices into consideration. This dissertation will aim to address this research gap.

4.2 Hypothesis development

When it comes to spending behaviour, the existing literature shows that the famine survivors tend to have higher savings in banks (Chen et al., 2018) and on average waste food less (Ding et al., 2022), relative to those who were not affected by famine. Additionally, they tend to spend less on entertainment and travel (Yao et al., 2020). Taken together, famine survivors on average appear to spend less and demonstrate frugal behaviour. To put in the framework of this dissertation, famine survivors on average have lower food expenditure.

Previous *Sections* confirm this position. There, I show that famine survivors are in the "earlier phase of nutrition transition" comparing to famine non-exposed counterparts. This earlier phase is characterized by higher share of cereals intake relative to later phases of nutrition transition. Not only that the cereals, and in case of China it is rice, are very affordable, but they are also the main source of carbohydrates in traditional Asian diets (Han et al., 2013; Rebello et al., 2014). Therefore, higher consumption of rice of famine survivors, explains higher intake of carbohydrates and lays foundation for my hypothesis:

H1: Famine survivors spend less on food items, relative to those who were not affected by famine

Literature which explores famine induced death rates, include in the assessment not only those who died directly because of famine related factors, but it also includes those which for some reason were not born because a famine occurred (Meng et al., 2015). The Great Famine claim the highest number of lives in recorded history, and as such it had extraordinary consequences for those who survived it. In the previous two *Sections* I find those consequences to be related to Dietary Diversity Score and nutrient intake. Here, I argue that:

H2: Famine survivors postpone giving birth at a higher rate, relative to those who were not affected by famine

4.3 Data

In this study I use China Health and Retirement Longitudinal Study (CHARLS) as the main data source. The dataset is publicly available and requires registration prior to data access. CHARLS is a panel survey of people aged 45 and over and their partners regardless of age in China. "It was designed to better understand the socioeconomic determinants and consequences of aging. The survey includes a rich set of questions regarding economic standing, physical and psychological health, demographics, and social networks of aged persons. The survey is designed to ensure comparability with the Health and Retirement Survey (HRS) in the United States and related aging surveys, such as the English Longitudinal Study of Aging (ELSA) in England and the Survey of Health, Aging and Retirement in Europe (SHARE) in Europe and Israel." (Harmonized CHARLS Documentation, July 2021)

The main goal of this longitudinal dataset is to provide a high quality nationally representative sample of Chinese residents' data to serve the needs of scientific research on health, economic position, and quality of life as people age. The survey elicits information about demographics, income, assets, health, cognition, family structure and connections, health care use and costs, housing, job status and history, expectations, biomarkers, and insurance.

"The first wave of CHARLS was conducted between June 2011 and March 2012. The sample population was selected as part of a stratified, multistage probability design. The first component of this sampling framework was the probability proportion to size (PPS) sampling of all county-level unit except for Tibet after stratifying by region, characteristic of the county (urban or rural), and per-capita gross domestic product (GDP). Households were selected within PSUs using a CHARLS-designed mapping/listing software (CHARLS-GIS) that uses Google Earth images to list all dwelling units in all residential buildings to create sampling frames. If the sampled household had occupants older than 40, one of them was randomly selected. If the selected person was aged 45 or older, they became a respondent. If the selected person was between age 40 and 44, they were reserved as a refreshment sample. This initial sample included 17,708 respondents in 10,257 households in 450 villages/urban communities in 150 counties/districts in 28 provinces. The second wave was conducted between July 2013 and January 2014 and included a refreshment sample consisting of individuals aged between 43 and 44 at Wave 1 and their partners. The third wave was conducted between July 2015 and January 2016 and included a refreshment sample consisting of individuals aged between 41 and 42 at Wave 1 and their partners. The fourth wave was conducted between July and November 2018 and included a refreshment sample consisting of individuals who were 40-

years-old at Wave 1 and their partners. The data include any individual interviewed at least once. This includes respondents and current and former spouses regardless of age." (Harmonized CHARLS Documentation, July 2021)

The Harmonized CHARLS dataset used in the analysis incorporates the demographic background data, family information data, family transfer data, health care and insurance data, health status and function data, household income data, household roster data, housing characteristics data, individual income data, weight data, and work, retirement and pension data. It does not include any data which is not for public release. The main data questionaries used in constructing CHARLS harmonized data files are presented below:

1. Demographic information (respondent and spouse): place and date of birth; residence and migration; Hukou information; Education; Marital status.
2. Household roster (nonrespondent household member): Demographic information; Relationship with the main respondent.
3. Family: All parents and children (demographic, occupation); Siblings (simple aggregate information); Interactions of each family member (time-spent together, two-way financial exchanges).
4. Health status and functioning: Self-reported general health; Doctor diagnosed diseases; Lifestyle and life behavior: Functional limitations and helpers; Cognition.
5. Health care and Insurance: Current and past insurance; Health care utilization: Health care costs.
6. Work, retirement and pension: Current job status; Work history; Current and most recent job; Unemployment; Retirement; Pension.
7. Income, expenditures and assets: Household income and expenditures; Household assets; Individual assets.
8. Housing characteristics: Construction material; home facilities; Cleanliness and temperature.
9. Interviewer observations

Source: Zhao et al. (2014)

In addition to the four waves, I rely on the CHARLS Life History Survey conducted on the same respondents in 2014 for information on their life history (CHARLS Life History Survey). This survey, collected by the CHARLS team, is a retrospective survey documenting events in migration, family, health, education, and employment of the CHARLS respondents since they were born. To ensure the accuracy of answers, the survey uses the event-history calendar method: respondents were first reminded of the timing of important national events (such as the Great Famine of China), then, as personal history unfolded, important events in the respondents' residential and family histories were used to anchor answers to those histories. In this way, the survey successfully recorded all respondents' firsthand answers to important questions related to their prior experience to adverse events, such as famine. Using unique respondent (and household) IDs, I am able to combine the CHARLS Life History Survey data set with CHARLS Harmonized data files spanning the 2011, 2013, 2015 and 2018 wave, resulting in a panel dataset comprised of 30,310 observations, tracking 15,199 unique individuals over the 4 survey waves.

4.4 Methodology and empirical strategy

Methodology

Unlike CHNS dataset, where I had to introduce a proxy measure for exposure to famine (EDR), CHARLS dataset provides explicit information about food insecurity episodes and the Great Famine. The respondents answered to the three questions:

- When you were a child before age 17 was there ever a time when your family did not have enough food to eat?
- Between 1958-1962[7] did you and your family (including your grandparents, parents, siblings, children and so on) experience starvation?
- During those days, had any of your family (including your grandparents, parents, siblings, children and so on) starved to death?

I translated these three questions into main explanatory variables.

There is also a limitation to this approach.

All three questions provide yes/no answers.

In Question 1, my analysis would benefit from deeper understanding of how respondent defines "enough food", and if it was only one family member who had to skip a meal or reduce regular food intake, or it was whole family who sacrificed. Perception of these two dimensions might have differed among respondents.

In Question 2, by applying this binary answer approach, I assume that, in the analysis of effects of starvation during the Great Famine on food expenditures, there is no difference whether respondents were starving themselves or it was their grandparents who starved, as the dataset does not provide such information.

In case of Question 3, there is a follow-up question that probes how many family members died due to starvation during the Great Famine. As in previous case, if a respond stated that one family member died due to starvation, I had to assume that there is no difference if that one person was respondent's child or grandparent.

[7] Authors of CHARLS survey consider the Great Famine to occur in the period 1958-62; Please note that most authors agree that the famine happened around 1959-1961, and this period was used in *Sections Two* and *Three*; Note that this discrepancy does not affect my estimations.

In addition to these three, other independent variables, which I use to explore long-term effects of famine and food security episodes are listed below.

Household total income (yuan)

As per CHARLS harmonized data document, "it presents the sum of all income at the household level including income from earning income, capital income, pension income, income from government transfers, other income and the total income from other household members".

Household net value of financial assets (yuan)

This variable is a difference between the financial assets (cash, deposits in financial institutions, stocks, bonds, savings and other assets) and debt which includes the value of the household's mortgage of primary residence. Detailed description of all items which are included in this variable can be found in CHARLS harmonized data document.

Education of respondent

This variable includes three categories:

1. Less than lower secondary
2. Upper secondary & vocational training
3. Tertiary

Other explanatory variables used in my estimations are:

- Marital status
- Age
- Gender
- Area of residence
- Household size

Summary statistics are reported in *Table 3.1 – Table 3.2* below.

Table 3. 1:Summary statistics of explanatory control variables

Age presents age of respondent in years, at the time of interview; Gender presents whether respondent is male (1) or female (0); Rural presents whether respondent lives in rural (1) or urban (0) area; HH size presents number of persons living in the household; Marital status presents whether respondent is married (1) or not (0); HH financial assets presents a difference between the financial assets and debt (yuan); HH total income presents sum of all income at the household level (yuan); Educ1 presents if respondent has less than lower secondary education; Educ2 presents if respondent has upper secondary and vocational training; Educ3 presents if respondents has tertiary education.

Variables	mean	sd	min	max	N
Age	60.61	9.83	45	105	30310
Gender	0.47	0.50	0	1	30310
Rural	0.80	0.40	0	1	30310
HH size	3.11	1.60	1	16	30310
Marital status	0.86	0.34	0	1	30310
HH financial assets	478472.70	4481183.00	-4247550	2.70E+08	30310
HH total income	36260.83	77161.34	0	2626400	30310
Educ1	0.91	0.29	0	1	30310
Educ2	0.08	0.27	0	1	30310
Educ3	0.01	0.11	0	1	30310

In my estimations, sample size is 30310, and is involves 4 waves 2011, 2013, 2015 and 2018. Furthermore, average *age of respondent is 60.61*. In the sample 47% are males and 80% of the respondents are in rural areas. Average household size is 3.11 and 86% of the sample are married. Average household net financial assets are 478472.70 yuan, and average monthly household income is 36260 yuan. While 91% of the sample has less than lower secondary education, 8% has upper secondary and vocational training while 1% holds degree form tertiary education institution.

Table 3. 2: Summary statistics of key explanatory variables

GF experience presents whether the respondent and his family experienced starvation during the Great Famine (1) or not (0); Starved to death presents whether any family member starved to death during the Great Famine (1) or not (0); No food 0-17 presents whether respondent's family while he/she was a child before age 17 did not have enough to eat (1) or that has never happened (0); the following three variables present respondent's age in which their family had not enough to eat (1).

stats	mean	sd	min	max	N
GF experience	0.78	0.42	0	1	30310
Starved to death	0.11	0.32	0	1	30310
No food 0-17	0.70	0.46	0	1	30310
No food 0-5	0.35	0.48	0	1	30310
No food 6-12	0.53	0.50	0	1	30310
No food 13-17	0.37	0.48	0	1	30310

In the sample, 78% of respondents claim that they experienced starvation during the Great Famine, and 11% of those lost one or more of their family members to starvation during the Great Famine. 70% of the respondents claims that before they turned 17, at some point their family did not have enough to eat. Of those, 35% of respondents was 0-5 when their family experienced food shortage, 53% were 6-2, and 37% were between 13 and 17 years old.

Dependent variables in this chapter can be divided in three groups:

1. Variables which directly measure impact of famine and food insecurity episodes on nutrition
2. Variables which indirectly measure impact of famine and food insecurity episodes on nutrition
3. Variables which verify long lasting impact of exposure to famine

In the **first group** of variables, there are:

- Food expenditures (excluding expenditures of eating away from home)
- Eating out expenditures
- Household total food consumption expenditures
- Alcohol and tobacco expenditures

In the **second group**, there are family planning decision variables derived from the following question:

"*During those days [during the Great Famine], did the food shortage resulted in any of the following for your family (including your grandparents, parents, siblings, children and yourself)?*

- *Put off marriage*
- *Put off giving birth*
- *Could not giving birth*
- *Artificial abortion"*

There is a twofold rationale for including those dependent variables. The first one is driven by the existing literature which examines factors associated to dietary diversity score, and which found that there is inverse relationship between DDS of children 6-23 months of age and mother's age (Dangura & Gebremedhin, 2017). Each of the four famine-induced results might cause increased mother's age, hence the lower DDS of their children.

The second rationale is the intention to verify the "Excess Death Rate" literature, which argues that famine induced death rates, include not only those who died directly because of famine related factors, but it also includes those which for some reason were not born due to famine. Each of the four famine-induced results might affect the excess death rate.

In the **third group**, the following variables serve to verify long lasting impact of exposure to famine:

- *Entertainment expenditures*
- *Travel expenditures*
- *Donations to public expenditures*

All three categories of expenditures belong to non-essential expenditures. On one hand, I include those expenditures to test the frugal behaviour of those affected by famine, and on the other I test the existence of famine "echo-effect".

Summary statistics are reported in *Table 3.3 and Table 3.4* below

Table 3. 3: Summary statistics of key dependent variables

Food excl. eating out represents 7-day average household food expenditures which exclude eating out expenditures (yuan); Alcohol and tobacco represents 7-day average household expenditures on alcohol and tobacco (yuan); Eating out represents 7-day average household expenditures on food eaten away from home (yuan); HH total food represents 7-day average household expenditures on purchased food and food eaten from own production, meals eaten out as well as on alcohol and tobacco (yuan)

Household expenditures	mean	sd	min	max	N
Food excl. eating out	198.72	304.08	0	25,000	29308
Alcohol and tobacco	48.12	172.28	0	10,000	29259
Eating out	23.68	159.45	0	10,000	29390
HH total food	326.20	544.35	0	35,000	29259

On average, households from the sample spend 198.72 yuan on food expenditure weekly, and that excludes expenditures towards food eaten away from home. For alcohol and tobacco, households on average spend 48.12 yuan weekly, while this amount for eating out is 23.68 yuan. Total average food expenditures of the household in the past 7 days from is 326 yuan.

Table 3. 4: Summary statistics of other dependent variables

Put off marriage represents whether respondent or someone in respondents family put off marriage due to food shortage during the Great Famine (1) or not (0); Put off birth represents whether respondent or someone in respondents family put off birth due to food shortage during the Great Famine (1) or not (0); Couldn't give birth represents whether respondent or someone in respondents family could not give birth due to food shortage during the Great Famine (1) or not (0); Artificial abortion represents whether respondent or someone in respondents family decided to induce artificial abortion due to food shortage during the Great Famine (1) or not (0).

Family planning	mean	sd	min	max	N
Put off marriage	0.03	0.17	0	1	30310
Put off birth	0.03	0.16	0	1	30310
Couldn't give birth	0.02	0.14	0	1	30310
Artificial abortion	0.01	0.10	0	1	30310

Food shortage during the Great Famine caused certain respondents or their family member to make some family planning decisions. 3% of the sample put off their marriage during the Great Famine, 3%

of the sample put off their birth, 2% could not give birth and 1% of the sample decided to induce artificial abortion.

Empirical strategy

4.4.1 Effects of famine on lifestyle choices

To ascertain the effect of prior exposure to famine on individual's nutrition, explored through lifestyle choices later in life, I begin my analysis by estimating the following OLS regression on the panel dataset that follows 15,199 distinct individuals over four survey waves: 2011, 2013, 2015 and 2018:

$$LC_{i,h,j,t} = \beta_1 Famine\_Exp_{i,h,j} + \sum_{v=1}^{V} \beta_v X_{v,i,h,j,t} + \delta_t + \pi_j + \varepsilon_{i,h,j,t} \quad (1)$$

Where $LC_{i,h,j,t}$ denotes lifestyle choices of individual *i*, from household *h* living in province *j* at time *t*. I explore the following set of dependent variables: $LC_{i,h,t}$ = {*Food expenditure, Alcohol and tobacco expenditure, entertainment expenditures, (non-commute) travel expenditure, charitable donations, eating out expenditures, total household food consumption*}. Vector $X_{v,i,h,t}$ contains a set of *v* time-varying individual and household characteristics: marital status of the individual, age, gender, education, whether they live in an urban or rural area, household size and total income, and household net value of total financial assets. I also include year fixed effects δ_t and province fixed effects π_j, to control for time- and location-specific unobservable factors that can be driving the relation of interest. In all regressions I report heteroscedasticity-adjusted robust standard errors. I construct the main independent variable of interest *Famine_Exp* by using the 2014 Life History Survey, which explicitly asks respondents about their prior exposure to famine. Specifically, the 2014 Life History Survey asks the following set of questions:

1) Between 1958-1962 did you and your family (including your grandparents, parents, siblings, children and so on) experience starvation?
2) During those days, had any of your family (including your grandparents, parents, siblings, children and so on) starved to death?
3) When you were a child before age 17 was there ever a time when your family did not have enough food to eat?
    a. At what age ranges did this (your family had not enough food to eat) happen?

Using the information from question 1) above, I construct my first measure of *Famine_Exp* , which I denote *GF experience. GF experience* equals 1 if respondent's family starved during the Great Chinese Famine 1958-1962, and 0 otherwise:

$$GF\ experience = \begin{bmatrix} 1 & if\ respondent's\ family\ starved\ during\ the\ Great\ Famine \\ 0 & otherwise \end{bmatrix}$$

In my analysis, I also use two additional measures of exposure, that utilize information from question 2 and 3 above: $No\ food\ 0-17$ and $Starved\ to\ Death$. $No\ food\ 0-17$ equals 1 if respondent did not have enough food at some time point during their childhood, before age 17 and 0 otherwise:

$$No\ food\ 0-17 = \begin{bmatrix} 1 & if\ respondent\ did\ not\ have\ enough\ food \\ 0 & otherwise \end{bmatrix}$$

$Starved\ to\ Death$ equals 1 if there was someone in respondent's household who starved to death during the Great Chinese Famine 1958-62, and 0 otherwise:

$$Starved\ to\ Death = \begin{bmatrix} 1 & if\ in\ respondent's\ household\ someone\ starved\ to\ death\ during\ Great\ Famine \\ 0 & otherwise \end{bmatrix}$$

To ascertain the effect of prior exposure to famine on individual's family planning decisions later in life, and for the seasons elaborated in 5.1 "Methodology" section above, I conduct my analysis by estimating the following panel OLS regression:

$$FP_{i,h,j,t} = \beta_1 Famine\_Exp_{i,h,j} + \sum_{v=1}^{V} \beta_i X_{v,i,h,j,t} + \delta_t + \pi_j + \varepsilon_{i,h,j,t} \quad (2)$$

Where $FP_{i,h,j,t}$ denotes family planning decisions of individual *i*, from household *h*, located at province *j* at time *t*. I explore the following set of dependent variables: $FP_{i,h,j,t}$ = {*Put off marriage, put off giving birth, couldn't give birth, artificial abortion*}. The other variables are defined as above.

Since the dependent variables $FP_{i,h,t}$ are dichotomous variables, I estimate Equation (2) above as a linear probability model (LPM), which allows me to include year and province fixed effects, δ_t and π_j.

4.4.2 Heterogenous effects of famine exposure by age

Motivated by results from earlier chapters, I utilize the wealth of information about famine exposure available in the 2014 Life History Survey, to analyse the importance of *timing* of exposure to famine. In particular, I am interested in seeing how the average results obtained by estimating Equation (1) vary with the *age* at which the respondent was exposed to food shortage. Specifically, I estimate the following specification:

$$LC_{i,h,j,t} = \beta_1 Nofood0\_17_{i,h,j} + \sum_{v=1}^{V} \beta_i X_{v,i,h,j,t} + \delta_t + \pi_j + \varepsilon_{i,h,j,t} \quad (4)$$

Where $Nofood0\_17$ denotes age when the respondent's family did not have enough food to eat, and which is derived from Question 3, above, to denote the age range between 0 and 5 (6-12, and 13-17 respectively) when the respondent or someone in their household did not have enough food to eat.

4.5 Results

4.5.1 Impact of early-life exposure to famine on life-style choices later in life

To ascertain the effect of prior exposure to famine on individual's lifestyle choices later in life, I begin my analysis by estimating Equation (1). Results are reported in *Table 3.5*.

Panel A shows the results that include year fixed effects, and Panel B shows the results that include both year and province fixed effects. I report heteroscedasticity-adjusted robust standard errors.[8] The main independent variables of interest *GF experience* equals 1 if respondent's family starved during the Great Chinese Famine 1958-1962, and 0 otherwise:

$$GF\ experience = \begin{bmatrix} 1 & if\ respondent's\ family\ starved\ during\ the\ Great\ Famine \\ 0 & otherwise \end{bmatrix}$$

As we can see from Column 1 in Panel A, the estimated coefficient on *GF experience* is negative and statistically significant at 1%, suggesting that respondents' whose family members have been starved in the past tend to have lower Food Expenditures (excluding spending on eating out). Dependent variable in Column 2 is Alcohol and Tobacco Expenditure: the estimated coefficient on *Famine* is again negative, yet statistically significant at 10%, suggesting that prior exposure to famine lowers respondents spending on alcohol and tobacco later in life. Column 3 reports the results for Entertainment related expenditures - the estimated coefficient on *GF experience* is again negative, but not statistically significant at conventional levels. In Column 4 the dependent variable is Travel-related expenditures, and the estimated coefficient on *GF experience* is again negative, and statistically significant at 5% level, suggesting that prior exposure to famine lowers respondents spending on non-work related travel. In Column 5 the dependent variable is charitable Donations - the estimated coefficient on *GF experience* is now positive, but not statistically significant, suggesting that prior exposure to famine increased respondents' charitable donations, although the

[8] Results remain quantitively similar when clustering standard errors by province.

relationship is not statistically meaningful. Column 6 reports coefficient estimates when Eating Out Expenditures are used as a dependent variable.

Table 3. 5: Effects of starvation during the Great Famine on household expenditures; Year FE
Explanatory Variables are defined as described in Table 3.1; GF experience presents whether the respondent and his family experienced starvation during the Great Famine (1) or not (0);
Dependent variables are defined as follows Food excl. eating out represents household food expenditures in the past 7 days which exclude eating out expenditures (yuan); Alcohol and tobacco represents household expenditures in the past 7 days on alcohol and tobacco (yuan); Eating out represents household expenditures in the past 7 days on food eaten away from home (yuan); HH total food represents household expenditures in the past 7 days on purchased food and food eaten from own production, meals eaten out as well as on alcohol and tobacco (yuan); Entertain., Travel and Donations represents expenditures of a household in the past 30 days for entertainment purposes, non-work related travel and charitable donations.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include year fixed effects.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A	Food excl. eating out	Alcohol and tobacco	Entertain.	Travel	Donations	Eating Out	HH total food
GF experience	-18.711***	-5.208*	-2.725	-180.781**	16.358	-3.163	-30.048***
	[-3.876]	[-1.909]	[-0.480]	[-2.130]	[1.075]	[-1.225]	[-3.159]
Marital status	9.968	3.479	1.979	45.002	9.137	-2.949*	18.248
	[1.109]	[0.944]	[0.222]	[0.313]	[1.023]	[-1.705]	[1.351]
Age	-1.088***	-1.174***	0.485**	2.57	0.034	-0.862***	-3.718***
	[-4.930]	[-10.072]	[2.279]	[0.951]	[0.062]	[-8.925]	[-9.507]
Gender	-0.342	5.257***	-3.149	-93.543**	-2.912	1.298	8.079
	[-0.109]	[2.926]	[-0.724]	[-2.349]	[-0.211]	[0.776]	[1.377]
Rural	-101.827***	-2.601	-19.753***	-708.030***	-21.717	-19.204***	-72.771***
	[-22.985]	[-0.857]	[-5.786]	[-10.894]	[-1.453]	[-6.019]	[-8.830]
Educ1	-100.093***	6.739	10.122	-3,440.341***	-41.281	-70.549***	-172.399***
	[-5.286]	[0.909]	[0.336]	[-4.511]	[-0.527]	[-4.484]	[-4.732]
Educ2	-56.318***	1.84	3.06	-2,953.450***	-34.637	-46.374***	-119.847***
	[-2.873]	[0.233]	[0.099]	[-3.823]	[-0.422]	[-2.838]	[-3.234]
HH total Income	0.273***	0.131***	0.067***	4.945***	1.610*	0.284***	0.678***
	[4.761]	[3.352]	[3.690]	[3.991]	[1.847]	[3.767]	[4.705]
HH size	24.359***	4.296***	0.609	-12.258	-6.157	0.774	36.184***
	[19.107]	[6.656]	[0.380]	[-0.939]	[-0.875]	[1.167]	[16.069]
HH financial assets	0.386***	-0.010	0.009**	13.279***	0.074	0.009	0.694*
	[3.574]	[-1.221]	[2.343]	[57.866]	[0.125]	[0.509]	[1.801]
Year FE	Y	Y	Y	Y	Y	Y	Y
Observations	29,257	29,208	29,339	29,208	29,420	29,339	29,208
R-squared	0.068	0.016	0.005	0.113	0.018	0.045	0.065
AIC	415,492	393,071	438,910	588,325	502,832	390,470	429,830
BIC	415,583	393,162	439,002	588,416	502,923	390,562	429,921

The estimated coefficient on *GF experience* is again negative, yet not statistically significant at conventional levels, suggesting that prior exposure to famine lowers respondents spending on eating out, although the relation is not statistically significant. In Column 7, I report the estimates for Household total food consumption (as measured over the 7 days prior to interview date). The estimated coefficient on *GF experience* is again negative, and statistically significant at 1% level, suggesting that prior exposure to famine lowers respondents' household's total food consumption later in life. It is important to note that in these regressions I control for year fixed effects, which absorb any time-specific macro-economic factors that can be driving my results.

In Panel B (*Table 3.6*), in addition to year fixed effects, I also include province fixed effects, which control for time-invariant province level unobservable factors that can be driving my results. Qualitatively, the results remain unchanged, although some of the estimated coefficients on *GF experience* now lose significance (for example: on Alcohol and Tobacco spending, and on travel expenditures). In Column 1, the dependent variable is Food Expenditures (excluding spending on eating out). Estimated coefficient on *GF experience* remains negative and significant at 5% level, with a slightly lower magnitude than that reported in Column 1 Panel A. This suggests that omitted time-invariant factors at the province level did explain part of that negative relation, although the effect remains significant. In Column 7, the dependent variable is total Household Food Consumption. Estimated coefficient on *GF experience* remains negative and significant at 5% level, with a slightly lower magnitude than that reported in Column 7 of Panel A. Again, this indicates that omitted time-invariant factors at the province level did play a role, although the effect remains significant. To further investigate this, I test relative goodness-of-fit. AIC suggests that specifications which include both province and wave fixed effect seem to perform better relative to those that include only wave fixed effect. However, based on BIC, specifications which include only wave fixed effect seem to exhibit better relative goodness-of-fit comparing to those that include both province and wave fixed effect.

Table 3. 6: Effects of starvation during the Great Famine on household expenditures; Year/Province FE
Explanatory Variables are defined as described in Table 3.1; GF experience presents whether the respondent and his family experienced starvation during the Great Famine (1) or not (0);
Dependent variables are defined as follows Food excl. eating out represents household food expenditures in the past 7 days which exclude eating out expenditures (yuan); Alcohol and tobacco represents household expenditures in the past 7 days on alcohol and tobacco (yuan); Eating out represents household expenditures in the past 7 days on food eaten away from home (yuan); HH total food represents household expenditures in the past 7 days on purchased food and food eaten from own production, meals eaten out as well as on alcohol and tobacco (yuan); Entertain., Travel and Donations represents expenditures of a household in the past 30 days for entertainment purposes, non-work related travel and charitable donations.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include year and province fixed effects.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel B	Food excl. eating out	Alcohol and tobacco	Entertain.	Travel	Donations	Eating Out	HH total food
GF experience	-12.720**	-4.038	-2.381	-115.239	20.225	-2.142	-22.016**
	[-2.568]	[-1.444]	[-0.431]	[-1.392]	[1.361]	[-0.853]	[-2.260]
Married	10.395	3.353	5.611	47.059	17.437**	-3.412**	18.993
	[1.159]	[0.908]	[0.631]	[0.336]	[2.159]	[-1.967]	[1.405]
Age	-1.301***	-1.177***	0.648***	1.184	0.208	-0.846***	-3.852***
	[-5.720]	[-10.305]	[2.812]	[0.493]	[0.441]	[-8.332]	[-9.670]
Gender	-0.309	4.917***	-3.489	-91.778**	-4.603	1.199	6.964
	[-0.099]	[2.733]	[-0.806]	[-2.263]	[-0.333]	[0.718]	[1.192]
Rural	-101.236***	-5.157	-17.885***	-665.708***	-32.448**	-18.165***	-73.261***
	[-22.482]	[-1.624]	[-5.000]	[-10.833]	[-2.492]	[-5.810]	[-8.763]
Educ1	-98.770***	4.076	12.475	-3,266.776***	-55.539	-70.009***	-174.980***
	[-5.356]	[0.553]	[0.405]	[-4.313]	[-0.756]	[-4.439]	[-4.821]
Educ2	-53.484***	1.266	3.410	-2,760.003***	-42.641	-45.824***	-115.818***
	[-2.796]	[0.161]	[0.108]	[-3.601]	[-0.526]	[-2.801]	[-3.135]
HH Total Income	0.252***	0.126***	0.082***	4.700***	1.631*	0.280***	0.651***
	[4.540]	[3.231]	[3.596]	[3.780]	[1.830]	[3.684]	[4.576]
HH size	24.224***	4.360***	-0.230	-19.712	-9.712	1.070	35.822***
	[18.364]	[6.522]	[-0.143]	[-1.485]	[-1.113]	[1.455]	[15.309]
HH financial assets	0.357***	-0.012	0.008*	13.173***	0.102	0.007	0.677*
	[3.297]	[-1.489]	[1.660]	[64.099]	[0.162]	[0.395]	[1.737]
Year FE	Y	Y	Y	Y	Y	Y	Y
Province FE	Y	Y	Y	Y	Y	Y	Y
Observations	29,257	29,208	29,339	29,208	29,420	29,339	29,208
R-squared	0.078	0.021	0.014	0.119	0.020	0.047	0.073
AIC	415,230	392,981	438,690	588,154	502,820	390,440	429,621
BIC	415,545	393,296	439,006	588,470	503,136	390,756	429,934

Table 3.7 presents coefficients of regional dummy variables. As we can see, there is a substantial variation in regional unobservable characteristics relative to the reference province – Yunnan. Interestingly, those provinces which were most affected by the Great Famine have negative coefficients, while the provinces which were least affected have positive coefficients. Also, it must be

acknowledged that the panel dataset, and the associated unobservable time invariant regional characteristics refer to the survey period (2011-2018) rather than the period when the Great Famine occurred (1958-1961).

Table 3. 7 Coefficient of regional dummy variables, which capture unobservable regional characteristics. Dependent variables are defined as in table 3.5.

	Food excl. eating out	Alcohol and tobacco	Entertain.	Travel	Donations	Eating Out	HH total food
Fujian	70.667***	21.967***	-17.983	-316.786	-40.374	-6.783*	50.659**
	[6.784]	[2.703]	[-1.119]	[-1.295]	[-0.571]	[-1.746]	[2.531]
Qinghai	-18.054	-13.524**	10.411***	-243.037	-69.272	-3.333	-62.284***
	[-1.249]	[-2.352]	[3.716]	[-1.007]	[-1.142]	[-0.851]	[-3.100]
Sichuan	31.144***	-5.020	-14.239**	-280.478	-83.771	4.293	21.073
	[2.763]	[-0.877]	[-2.093]	[-1.224]	[-1.295]	[1.190]	[1.006]
Hebei	-6.486	-20.215***	-7.450	-498.727**	-98.663	3.104	-53.870***
	[-0.825]	[-5.049]	[-1.050]	[-2.119]	[-1.306]	[0.668]	[-3.505]
Jiangxi	31.893***	-16.529***	-0.998	-145.245	-88.887	-14.488***	-15.059
	[3.628]	[-4.806]	[-0.063]	[-0.580]	[-1.219]	[-4.095]	[-0.902]
Xinjiang	45.059**	-47.650***	-10.912***	-861.998***	-78.654	37.017***	-19.183
	[2.193]	[-9.617]	[-2.584]	[-2.683]	[-1.036]	[3.115]	[-0.688]
Beijing	148.139***	-41.679***	23.078	5,750.772***	-277.359	35.012	70.656
	[3.122]	[-5.432]	[1.520]	[2.664]	[-1.565]	[1.093]	[1.193]
Inner Mongolia	45.801***	-5.340	-3.980	-348.636	-101.311	12.992**	47.478**
	[4.035]	[-1.043]	[-1.566]	[-1.420]	[-1.246]	[2.195]	[2.176]
Jiangsu	33.913***	1.479	-175.469***	-767.275***	-252.572***	-3.907	-9.393
	[4.013]	[0.332]	[-4.576]	[-3.248]	[-2.671]	[-0.786]	[-0.604]
Chongqing	37.386	-12.139**	-6.916***	-666.629***	-111.675	-2.626	3.539
	[1.184]	[-2.572]	[-3.056]	[-3.093]	[-1.610]	[-0.518]	[0.098]
Gansu	-26.382***	-4.685	-10.942	-236.963	-91.681	11.941**	-54.612***
	[-3.897]	[-0.462]	[-0.943]	[-0.983]	[-1.392]	[2.394]	[-3.264]
Heilongjiang	-22.992**	-22.867***	-0.090	-156.767	-146.953*	-3.974	-88.056***
	[-2.142]	[-5.198]	[-0.014]	[-0.487]	[-1.794]	[-0.564]	[-5.192]
Guangdong	65.352***	-24.948***	3.162	-431.325*	-24.995	-3.666	-3.451
	[6.847]	[-6.780]	[1.348]	[-1.747]	[-0.308]	[-0.486]	[-0.187]
Liaoning	-14.064*	-17.428***	-3.827	-543.041**	-110.733	-2.380	-84.549***
	[-1.928]	[-5.128]	[-1.152]	[-2.354]	[-1.579]	[-0.543]	[-6.704]
Shannxi	-10.591	-1.993	-0.194	-553.109**	-115.460	0.005	-61.427**
	[-0.825]	[-0.155]	[-0.103]	[-2.365]	[-1.614]	[0.001]	[-2.380]
Shanghai	124.356***	5.767	3.084	63.952	-272.434	33.084	125.568
	[3.126]	[0.284]	[0.206]	[0.119]	[-1.450]	[0.705]	[1.524]
Tianjin	129.881***	-5.065	-1.885	417.977	-105.294	17.977	60.967**
	[5.715]	[-0.711]	[-0.531]	[0.939]	[-1.192]	[1.479]	[1.988]
Zhejiang	54.873***	12.840**	-4.604	450.272	-42.820	22.275	64.981**
	[6.279]	[2.052]	[-0.336]	[1.317]	[-0.344]	[1.470]	[2.379]
Jilin	-23.270***	-20.151***	-9.070***	-348.979	-108.780	6.859	-82.589***
	[-2.798]	[-4.640]	[-3.941]	[-1.388]	[-1.484]	[0.993]	[-5.381]
Guangxi	16.268*	-22.288***	10.495	-246.341	-89.999	-7.038	-61.806***
	[1.926]	[-5.814]	[0.908]	[-0.997]	[-1.277]	[-1.548]	[-4.267]
Anhui	13.026*	-3.776	-16.809*	-450.513**	-76.473	2.458	-26.771*
	[1.710]	[-0.948]	[-1.889]	[-2.082]	[-1.148]	[0.687]	[-1.904]
Hubei	6.095	-7.681*	-0.646	-543.890**	-107.393	-0.829	-3.184
	[0.678]	[-1.831]	[-0.352]	[-2.425]	[-1.536]	[-0.202]	[-0.200]
Shanxi	-7.408	-4.340	-0.671	-318.930	-91.725	-0.137	-27.778
	[-0.829]	[-0.477]	[-0.266]	[-1.294]	[-1.232]	[-0.032]	[-1.418]
Shandong	-20.042***	-25.993***	-4.380**	-586.702***	-111.590	-6.713**	-107.953***
	[-3.183]	[-8.487]	[-2.089]	[-2.581]	[-1.463]	[-1.975]	[-9.298]
Henan	-28.490***	-24.659***	-2.392*	-381.366*	-94.942	-3.641	-103.374***
	[-4.808]	[-7.418]	[-1.730]	[-1.678]	[-1.389]	[-1.053]	[-8.419]
Hunan	-11.178	-6.428	-2.341	-362.198	-107.071	6.573	-31.259**
	[-1.529]	[-1.285]	[-1.177]	[-1.536]	[-1.289]	[1.051]	[-2.026]
Guizhou	-6.884	-2.494	-3.866	-421.321*	-11.584	6.268	-35.200
	[-0.541]	[-0.207]	[-0.131]	[-1.801]	[-0.134]	[1.021]	[-1.487]

4.5.2 Impact of early-life famine on family planning later in life

Results presented in the previous sub-section suggest that early-life exposure to famine, as proxied by whether the respondent's household has responded affirmatively to survey question if they have been starved during the Great Famine, has significant long-lasting impact on respondent's lifestyle choices, in particular when it comes to food consumption, travel, eating out, entertainment and alcohol and tobacco spending, as well as on individual's propensity to donate to charitable purposes. In this subsection, I further study the long-lasting impact of famine, by extending the analysis to understand the effects on respondent's family planning decisions. In particular, I study how early life exposure to famine of respondents, or their family members affects the propensity of respondents to delay marriage and have children, as well the potential effect of famine on not being able to conceive or to artificially terminate unwanted pregnancies.

To determine the effect of prior exposure to famine on individual's family planning decisions later in life, I begin my analysis by estimating a linear probability model (LPM), as shown in Equation (2). Results are reported in *Table 3.8*.

Panel A shows the results that include year fixed effects, and Panel B shows the results that include both year and province fixed effects. The main independent variable of interest *GF experience* is defined as before. As we can see from Column 1 in Panel A, where the dependent variable is *Put off Marriage*, the estimated coefficient on *GF experience* is positive and statistically significant at 1% level, suggesting that respondents' whose family members have been starved in the past tend to be more likely to delay getting married, controlling for a host of individual and household characteristics. Dependent variable in Column 2 is *Put off baby*: the estimated coefficient on *GF experience* is again positive, and also statistically significant at 1% level, suggesting that prior exposure to famine increases the probability of delaying having children later in life. Column 3 reports the results for inability to have children as the dependent variable, and the estimated coefficient on *GF experience* is again positive, and also statistically significant at 1% level, suggesting that individuals whose families have been exposed to famine tend to be less likely to be able to have children later in life. In Column 4 the dependent variable is Artificial Abortion, and the estimated coefficient on *Famine* is again positive, and statistically significant at 1% level, suggesting that prior exposure to famine increases the probability of having an abortion later in life.

Table 3. 8: Effects of starvation during the Great Famine on family planning decisions; linear probability model; Year FE
Explanatory Variables are defined as described in Table 3.1; GF experience presents whether the respondent and his family experienced starvation during the Great Famine (1) or not (0);
Dependent variables are defined as described in Table 3.4
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include year fixed effects.*

	(1)	(2)	(3)	(4)
Panel A	Put off marriage	Put off birth	Couldn't give birth	Artificial abortion
GF experience	0.018***	0.018***	0.013***	0.006***
	[10.141]	[10.630]	[8.968]	[5.699]
Marital status	-0.011***	-0.008***	-0.002	-0.008***
	[-3.211]	[-2.591]	[-0.905]	[-3.816]
Age	0.000***	0.001***	0.001***	0.000**
	[3.655]	[5.787]	[7.650]	[2.510]
Gender	0.008***	0.003*	0.002	-0.001
	[3.885]	[1.781]	[1.060]	[-0.984]
Rural	0.011***	0.008***	0.008***	0.006***
	[4.973]	[3.785]	[4.873]	[5.634]
Educ1	-0.022**	-0.013	0.003	-0.016**
	[-1.974]	[-1.307]	[0.428]	[-2.051]
Educ2	-0.024**	-0.018*	-0.005	-0.017**
	[-2.060]	[-1.815]	[-0.778]	[-2.148]
HH Total Income	0.000	-0.000*	0.000	-0.000
	[0.087]	[-1.928]	[0.076]	[-1.625]
HH size	0.002***	0.001	0.001**	-0.000
	[2.837]	[1.236]	[2.399]	[-0.191]
HH financial assets	-0.000	-0.000	0.000	-0.000
	[-0.425]	[-1.476]	[0.074]	[-1.497]
Year FE	Y	Y	Y	Y
Observations	30,270	30,270	30,270	30,270
R-squared	0.005	0.006	0.006	0.003
AIC	-23,280	-26,008	-35,151	-55,900
BIC	-23,189	-25,917	-35,059	-55,808

In Panel B (*Table 3.9*), I repeat the analysis, but now including both year and province fixed effects. This ensures that I control for any time- and province-specific unobservable factors that can be driving my results. The results remain unchanged, both in terms of the economic magnitudes and statistical significance. As it can be seen from Panel B, respondents whose families have been starved in the past, tend to be more likely to delay getting married and having children, as indicated by positive and highly significant coefficient on *GF experience* (Columns 1 and 2). They are also less likely to have children, and more likely to have an abortion (Columns 3 and 4).

Table 3. 9: Effects of starvation during the Great Famine on family planning decisions; linear probability model; Year/Province FE Explanatory Variables are defined as described in Table 3.1; GF experience presents whether the respondent and his family experienced starvation during the Great Famine (1) or not (0);
Dependent variables are defined as described in Table 3.4
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include year and province fixed effects.*

	(1)	(2)	(3)	(4)
Panel B	Put off marriage	Put off birth	Couldn't give birth	Artificial abortion
GF experience	0.020***	0.018***	0.012***	0.006***
	[10.415]	[10.483]	[8.126]	[5.289]
Marital status	-0.008**	-0.006*	-0.001	-0.008***
	[-2.403]	[-1.808]	[-0.250]	[-3.509]
Age	0.000***	0.001***	0.001***	0.000**
	[3.459]	[5.581]	[8.600]	[2.453]
Gender	0.008***	0.003*	0.002	-0.001
	[3.819]	[1.713]	[0.944]	[-1.089]
Rural	0.007***	0.004**	0.006***	0.006***
	[3.179]	[1.984]	[3.573]	[5.444]
Educ1	-0.024**	-0.015	0.002	-0.016**
	[-2.117]	[-1.510]	[0.416]	[-2.103]
Educ2	-0.024**	-0.018*	-0.003	-0.016**
	[-2.096]	[-1.785]	[-0.476]	[-2.073]
HH Total Income	0.000	-0.000	0.000	-0.000
	[0.203]	[-1.391]	[0.593]	[-1.501]
HH size	0.001	-0.000	0.001	-0.000
	[1.278]	[-0.644]	[1.145]	[-0.947]
HH financial assets	-0.000	-0.000	0.000*	-0.000
	[-0.634]	[-1.281]	[1.683]	[-0.999]
Year FE	Y	Y	Y	Y
Province FE	Y	Y	Y	Y
Observations	30,270	30,270	30,270	30,270
R-squared	0.013	0.012	0.016	0.008
AIC	-23,474	-26,141	-35,411	-55,984
BIC	-23,158	-25,824	-35,095	-55,668

Table 3.10 presents coefficients of regional dummy variables. As we can see, similarly to the previous table which captures regional dummy variables, there is a variation in regional unobservable characteristics relative to the reference province – Yunnan. However, in this case, there is no overlap between the sign of the coefficient and the Great Famine severity.

Table 3. 10 Coefficient of regional dummy variables, which capture unobservable regional characteristics. Dependent variables are defined as in table 3.5.

	Put off marriage	Put off birth	Couldn't give birth	Artificial abortion
Fujian	0.040***	0.009	-0.008	-0.007
	[3.687]	[0.995]	[-1.129]	[-1.446]
Qinghai	0.057***	0.048***	0.096***	-0.007
	[3.375]	[2.927]	[4.913]	[-1.049]
Sichuan	0.009*	0.002	0.000	-0.001
	[1.709]	[0.287]	[0.032]	[-0.265]
Hebei	-0.001	-0.018***	-0.023***	-0.014***
	[-0.132]	[-3.380]	[-5.501]	[-4.471]
Jiangxi	0.008	-0.009	-0.014***	0.005
	[1.261]	[-1.460]	[-2.860]	[0.989]
Xinjiang	-0.011	-0.001	0.011	-0.006
	[-1.211]	[-0.060]	[0.778]	[-0.998]
Beijing	-0.027***	-0.035***	-0.028***	-0.013***
	[-5.253]	[-7.036]	[-6.476]	[-3.870]
Inner Mongolia	-0.015***	-0.027***	-0.009*	-0.009***
	[-3.284]	[-5.892]	[-1.930]	[-2.822]
Jiangsu	-0.019***	-0.019***	-0.026***	-0.005
	[-3.998]	[-3.449]	[-6.107]	[-1.087]
Chongqing	-0.024***	-0.031***	-0.022***	-0.014***
	[-4.165]	[-5.344]	[-3.728]	[-4.067]
Gansu	0.009	-0.001	-0.003	0.000
	[1.258]	[-0.099]	[-0.401]	[0.067]
Heilongjiang	-0.018***	-0.029***	-0.019***	-0.010**
	[-3.127]	[-5.847]	[-3.769]	[-2.379]
Guangdong	0.011	-0.000	-0.009	-0.012***
	[1.596]	[-0.070]	[-1.563]	[-3.776]
Liaoning	-0.021***	-0.024***	-0.021***	-0.016***
	[-4.374]	[-4.601]	[-4.669]	[-6.004]
Shannxi	0.015**	-0.008	-0.005	0.003
	[2.023]	[-1.232]	[-0.800]	[0.522]
Shanghai	-0.022***	-0.028***	0.020	0.032
	[-4.641]	[-6.167]	[0.668]	[1.100]
Tianjin	-0.031***	-0.008	-0.031***	-0.017***
	[-8.028]	[-0.584]	[-8.250]	[-6.103]
Zhejiang	0.000	-0.017***	-0.027***	-0.010***
	[0.036]	[-3.078]	[-6.869]	[-2.979]
Jilin	-0.008	-0.012*	-0.015***	-0.003
	[-1.189]	[-1.697]	[-2.620]	[-0.584]
Guangxi	0.007	-0.002	-0.024***	-0.010**
	[0.909]	[-0.328]	[-5.133]	[-2.473]
Anhui	-0.005	-0.004	-0.006	-0.008**
	[-0.827]	[-0.674]	[-1.119]	[-2.204]
Hubei	-0.016***	-0.012*	-0.017***	-0.008*
	[-2.880]	[-1.721]	[-3.149]	[-1.877]
Shanxi	-0.016***	-0.027***	-0.021***	-0.013***
	[-3.090]	[-5.649]	[-4.909]	[-4.314]
Shandong	-0.014***	-0.024***	-0.016***	-0.012***
	[-3.130]	[-5.241]	[-3.775]	[-3.940]
Henan	-0.017***	-0.020***	-0.018***	-0.015***
	[-3.807]	[-4.315]	[-4.250]	[-5.485]
Hunan	0.010	-0.007	-0.006	0.003
	[1.453]	[-1.186]	[-1.084]	[0.579]
Guizhou	0.036***	0.016	-0.011	-0.001
	[2.629]	[1.297]	[-1.261]	[-0.094]

4.5.3 Alternative proxy for famine exposure: not having enough food

To ascertain whether the effect of prior exposure to famine on individual's lifestyle choices later in life reported above is sensitive to measurement issues, I use a different measure for famine exposure: $No\ food\ 0-17$. The main independent variables of interest $No\ food\ 0-17$ equals 1 if respondent did not have enough food at some time point during their childhood, and 0 otherwise:

$$No\ food\ 0-17 = \begin{bmatrix} 1 & if\ respondent\ did\ not\ have\ enough\ food \\ 0 & otherwise \end{bmatrix}$$

Using this alternative measure of exposure to famine, I re-estimate Equation (1). Results are reported in *Table 3.11.*

Panel A shows the results that include year fixed effects, and Panel B shows the results that include both year and province fixed effects. I report heteroscedasticity-adjusted robust standard errors.[9]

As we can see from Panel A, the estimated coefficient on *No Food 0-17* is negative across all specifications, but statistically significant at 1% level only in case of EatingOut expenditures. While the sign of the estimates is similar to those reported for *Famine* in *Table 3.5*, the lack of statistical significance points to the **important distinction between measuring exposure to famine using starvation (as it was the case for *GF Experience* versus lack of food** (as it is the case for *No Food 0-17).* Similar results are obtained in Panel B *(Table 3.12)*, with the inclusion of both year and province fixed effects.

[9] Results remain quantitively similar when clustering standard errors by province.

Table 3. 11: Effects of not having enough food before age 17 on household expenditures; Year FE

Explanatory Variables are defined as described in Table 3.1; No food 0-17 presents whether respondent's family while he/she was a child before age 17 did not have enough to eat (1) or that has never happened (0);
Dependent variables are defined as follows Food excl. eating out represents household food expenditures in the past 7 days which exclude eating out expenditures (yuan); Alcohol and tobacco represents household expenditures in the past 7 days on alcohol and tobacco (yuan); Eating out represents household expenditures in the past 7 days on food eaten away from home (yuan); HH total food represents household expenditures in the past 7 days on purchased food and food eaten from own production, meals eaten out as well as on alcohol and tobacco (yuan); Entertain., Travel and Donations represents expenditures of a household in the past 30 days for entertainment purposes, non-work related travel and charitable donations.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include year fixed effects.*

Panel A	(1) Food excl. eating out	(2) Alcohol and tobacco	(3) Entertain.	(4) Travel	(5) Donations	(6) Eating Out	(7) HH total food
No food 0-17	-3.290	-1.608	-2.606	-89.204	-9.227	-6.578***	-10.769
	[-0.865]	[-0.735]	[-0.566]	[-1.395]	[-0.527]	[-2.816]	[-1.552]
Marital status	9.358	3.334	1.980	41.450	10.199	-2.805	17.534
	[1.020]	[0.888]	[0.221]	[0.287]	[1.132]	[-1.630]	[1.275]
Age	-1.163***	-1.191***	0.487**	2.184	0.170	-0.839***	-3.804***
	[-4.781]	[-9.817]	[2.243]	[0.801]	[0.322]	[-9.039]	[-9.135]
Gender	-1.026	5.101***	-3.131	-97.092**	-1.644	1.509	7.261
	[-0.320]	[2.809]	[-0.728]	[-2.417]	[-0.114]	[0.885]	[1.225]
Rural	-102.285***	-2.687	-19.686***	-709.106***	-20.587	-18.903***	-73.178***
	[-23.278]	[-0.887]	[-5.795]	[-10.866]	[-1.383]	[-5.889]	[-8.917]
Educ1	-101.570***	6.430	10.265	-3,444.581***	-37.877	-69.817***	-173.917***
	[-5.356]	[0.868]	[0.338]	[-4.513]	[-0.478]	[-4.439]	[-4.778]
Educ2	-57.544***	1.558	3.097	-2,959.425***	-32.215	-46.006***	-121.309***
	[-2.929]	[0.197]	[0.099]	[-3.826]	[-0.392]	[-2.811]	[-3.272]
HH Total Income	0.275***	0.131***	0.067***	4.950***	1.607*	0.284***	0.680***
	[4.756]	[3.357]	[3.676]	[3.991]	[1.845]	[3.762]	[4.697]
HH size	24.337***	4.290***	0.613	-12.279	-6.092	0.792	36.152***
	[19.133]	[6.678]	[0.382]	[-0.940]	[-0.868]	[1.192]	[16.092]
HH financial assets	0.390***	-0.010	0.009**	13.285***	0.061	0.008	0.696*
	[3.597]	[-1.193]	[2.438]	[57.766]	[0.102]	[0.495]	[1.803]
Year FE	Y	Y	Y	Y	Y	Y	Y
Observations	29,257	29,208	29,339	29,208	29,420	29,339	29,208
R-squared	0.067	0.016	0.005	0.113	0.018	0.045	0.064
AIC	415,511	393,075	438,910	588,332	502,832	390,461	429,843
BIC	415,602	393,166	439,002	588,423	502,924	390,553	429,934

Table 3. 12: Effects of not having enough food before age 17 on household expenditures; Year and Province FE

Explanatory Variables are defined as described in Table 3.1; No food 0-17 presents whether respondent's family while he/she was a child before age 17 did not have enough to eat (1) or that has never happened (0);

Dependent variables are defined as follows Food excl. eating out represents household food expenditures in the past 7 days which exclude eating out expenditures (yuan); Alcohol and tobacco represents household expenditures in the past 7 days on alcohol and tobacco (yuan); Eating out represents household expenditures in the past 7 days on food eaten away from home (yuan); HH total food represents household expenditures in the past 7 days on purchased food and food eaten from own production, meals eaten out as well as on alcohol and tobacco (yuan); Entertain., Travel and Donations represents expenditures of a household in the past 30 days for entertainment purposes, non-work related travel and charitable donations.

*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include year and province fixed effects.*

Panel B	(1) Food excl. eating out	(2) Alcohol and tobacco	(3) Entertain.	(4) Travel	(5) Donations	(6) Eating Out	(7) HH total food
No food 0-17	-0.409	-1.780	-3.771	-70.265	-11.633	-6.315***	-10.116
	[-0.110]	[-0.836]	[-0.783]	[-1.081]	[-0.708]	[-2.930]	[-1.413]
Marital status	9.980	3.292	5.700	46.196	18.678**	-3.207*	18.723
	[1.092]	[0.875]	[0.636]	[0.330]	[2.243]	[-1.857]	[1.362]
Age	-1.367***	-1.188***	0.658***	0.987	0.385	-0.819***	-3.907***
	[-5.445]	[-9.962]	[2.781]	[0.406]	[0.887]	[-8.516]	[-9.194]
Gender	-0.858	4.827***	-3.395	-93.279**	-3.069	1.435	6.498
	[-0.268]	[2.653]	[-0.792]	[-2.285]	[-0.213]	[0.843]	[1.100]
Rural	-101.589***	-5.181	-17.755***	-665.540***	-31.257**	-17.894***	-73.396***
	[-22.711]	[-1.634]	[-4.958]	[-10.807]	[-2.421]	[-5.691]	[-8.804]
Educ1	-100.022***	3.928	12.823	-3,267.297***	-51.460	-69.265***	-175.727***
	[-5.420]	[0.533]	[0.413]	[-4.314]	[-0.692]	[-4.394]	[-4.846]
Educ2	-54.438***	1.112	3.585	-2,762.183***	-39.747	-45.416***	-116.617***
	[-2.843]	[0.142]	[0.114]	[-3.602]	[-0.490]	[-2.771]	[-3.157]
HH Total Income	0.252***	0.126***	0.082***	4.697***	1.629*	0.279***	0.651***
	[4.532]	[3.230]	[3.592]	[3.775]	[1.828]	[3.679]	[4.565]
HH size	24.191***	4.349***	-0.234	-19.971	-9.650	1.068	35.760***
	[18.382]	[6.527]	[-0.146]	[-1.508]	[-1.110]	[1.455]	[15.332]
HH financial assets	0.358***	-0.012	0.008*	13.174***	0.091	0.006	0.673*
	[3.304]	[-1.481]	[1.677]	[64.115]	[0.143]	[0.380]	[1.727]
Year FE	Y	Y	Y	Y	Y	Y	Y
Province FE	Y	Y	Y	Y	Y	Y	Y
Observations	29,257	29,208	29,339	29,208	29,420	29,339	29,208
R-squared	0.077	0.021	0.014	0.119	0.020	0.048	0.073
AIC	415,239	392,983	438,689	588,156	502,821	390,431	429,627
BIC	415,554	393,298	439,005	588,472	503,137	390,747	429,940

I next proceed to re-estimate Equation (2), to analyse the effect of prior exposure to famine on family planning decisions using this alternative exposure measure. Results are reported in *Table 3.13*. As before, Panel A shows the results that include year fixed effects, and Panel B shows the results that include both year and province fixed effects.

Table 3. 13: Effects of not having enough food before age 17 on family planning decisions; linear probability model; Year FE

Explanatory Variables are defined as described in Table 3.1; No food 0-17 presents whether respondent's family while he/she was a child before age 17 did not have enough to eat (1) or that has never happened (0)
Dependent variables are defined as described in Table 3.4
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include year fixed effects.*

	(1)	(2)	(3)	(4)
Panel A	Put off marriage	Put off birth	Couldn't give birth	Artificial abortion
No food 0-17	0.019***	0.014***	0.009***	0.009***
	[10.368]	[8.323]	[6.100]	[9.084]
Marital status	-0.011***	-0.008**	-0.002	-0.008***
	[-3.224]	[-2.560]	[-0.863]	[-3.873]
Age	0.000***	0.001***	0.001***	0.000**
	[3.520]	[5.794]	[7.660]	[2.184]
Gender	0.008***	0.003*	0.002	-0.001
	[3.795]	[1.788]	[1.105]	[-1.162]
Rural	0.011***	0.008***	0.008***	0.006***
	[4.745]	[3.652]	[4.767]	[5.351]
Educ1	-0.023**	-0.013	0.002	-0.017**
	[-2.064]	[-1.350]	[0.399]	[-2.151]
Educ2	-0.024**	-0.018*	-0.005	-0.017**
	[-2.082]	[-1.803]	[-0.756]	[-2.195]
HH Total Income	0.000	-0.000*	0.000	-0.000
	[0.182]	[-1.853]	[0.088]	[-1.404]
HH size	0.002***	0.001	0.001**	-0.000
	[2.789]	[1.205]	[2.375]	[-0.247]
HH financial assets	-0.000	-0.000	-0.000	-0.000
	[-0.345]	[-1.618]	[-0.147]	[-1.278]
Year FE	Y	Y	Y	Y
Observations	30,270	30,270	30,270	30,270
R-squared	0.006	0.005	0.005	0.004
AIC	-23,296	-25,996	-35,135	-55,930
BIC	-23,204	-25,905	-35,044	-55,838

The main independent variable of interest *No Food 0-17* is defined as before. As we can see from Column 1 in Panel A, where the dependent variable is *Put off Marriage*, the estimated coefficient on *No Food 0-17* is positive and statistically significant at 1% level, suggesting that respondents' who have experienced not having enough food in the past tend to be more likely to delay getting married, controlling for a host of individual and household characteristics. Dependent variable in Column 2 is *Put off baby*: the estimated coefficient on *No Food 0-17* is again positive, and also statistically significant at 1% level, suggesting that prior exposure to famine increases the probability of delaying having children later in life. Column 3 reports the results for inability to have children as the dependent variable, and the estimated coefficient on *No Food 0-17 is* again positive, and also statistically significant at 1% level, suggesting that individuals who at some point in their lives did not have enough food tend to be less likely to be able to have children later in life. In Column 4 the dependent variable is *Artificial Abortion* - the estimated coefficient on *No Food 0-17* is again positive, and statistically significant at 1% level, suggesting that prior exposure to famine increases the probability of having an abortion later in life.

In Panel B (*Table 3.14*), I repeat the analysis, but now including both year and province fixed effects. This ensures that I control for any time- and province-specific unobservable factors that can be driving my results. The results remain unchanged, both in terms of the economic magnitudes and statistical significance. As it can be seen from Panel B, respondents whose families have been starved in the past, tend to be more likely to delay getting married and having children, as indicated by positive and highly significant coefficient on *No Food 0-17* (Columns 1 and 2). They are also less likely to have children, and more likely to have an abortion (Columns 3 and 4).

Table 3. 14: Effects of not having enough food before age 17 on family planning decisions; linear probability model; Year and Province FE

Explanatory Variables are defined as described in Table 3.1; No food 0-17 presents whether respondent's family while he/she was a child before age 17 did not have enough to eat (1) or that has never happened (0)
Dependent variables are defined as described in Table 3.4
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include year and province fixed effects.*

	(1)	(2)	(3)	(4)
Panel B	Put off marriage	Put off birth	Couldn't give birth	Artificial abortion
No food 0-17	0.018***	0.013***	0.008***	0.008***
	[10.047]	[7.735]	[5.409]	[8.350]
Marital status	-0.008**	-0.006*	-0.001	-0.008***
	[-2.441]	[-1.805]	[-0.237]	[-3.578]
Age	0.000***	0.001***	0.001***	0.000**
	[3.447]	[5.697]	[8.655]	[2.197]
Gender	0.007***	0.003*	0.002	-0.001
	[3.778]	[1.756]	[0.999]	[-1.233]
Rural	0.007***	0.004*	0.006***	0.006***
	[3.022]	[1.902]	[3.511]	[5.242]
Educ1	-0.025**	-0.015	0.002	-0.017**
	[-2.177]	[-1.528]	[0.404]	[-2.184]
Educ2	-0.024**	-0.018*	-0.003	-0.016**
	[-2.104]	[-1.764]	[-0.451]	[-2.111]
HH Total Income	0.000	-0.000	0.000	-0.000
	[0.309]	[-1.273]	[0.627]	[-1.307]
HH size	0.001	-0.000	0.001	-0.000
	[1.338]	[-0.581]	[1.189]	[-0.919]
HH financial assets	-0.000	-0.000	0.000	-0.000
	[-0.486]	[-1.273]	[1.577]	[-0.753]
Year FE	Y	Y	Y	Y
Province FE	Y	Y	Y	Y
Observations	30,270	30,270	30,270	30,270
R-squared	0.014	0.011	0.015	0.008
AIC	-23,476	-26,121	-35,396	-56,007
BIC	-23,160	-25,804	-35,080	-55,691

4.5.3.1 Impact of "Not enough food" early in life on lifestyle choices and family planning later in life: heterogenous effects by age

Results presented in the previous chapters suggest that not all age groups are equally affected when exposed to famine. The previous subsections provide mixed evidence in this respect. On one hand, while not having enough food is negatively correlated to all food and non-food expenditures, it is statistically significant only for eating out expenditures (at 1%). On the other hand, *Age* is negatively correlated with all food expenditures (at 1%). This dichotomy inspired me to verify heterogenous effects of food shortage on different age groups (Equation 4).

Panel A (*Table 3.15*) shows the results that include year fixed effects, and Panel B (*Table 3.16*) shows the results that include both year and province fixed effects.

Table 3. 15: Effects of not having enough food in different age periods on household expenditures; Year FE

Explanatory Variables are defined as described in Table 3.1; No food 0-5 presents whether respondent's family while he/she was between 0 and 5 did not have enough to eat (1) or that has not happened (0); No food 6-12 presents whether respondent's family while he/she was between 6 and 12 did not have enough to eat (1) or that has not happened (0); No food 13-17 presents whether respondent's family while he/she was between 13 and 17 did not have enough to eat (1) or that has not happened (0); Dependent variables are defined as follows Food excl. eating out represents household food expenditures in the past 7 days which exclude eating out expenditures (yuan); Alcohol and tobacco represents household expenditures in the past 7 days on alcohol and tobacco (yuan); Eating out represents household expenditures in the past 7 days on food eaten away from home (yuan); HH total food represents household expenditures in the past 7 days on purchased food and food eaten from own production, meals eaten out as well as on alcohol and tobacco (yuan); Entertain., Travel and Donations represents expenditures of a household in the past 30 days for entertainment purposes, non-work related travel and charitable donations.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include year fixed effects.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A	Food excl. eating out	Alcohol and tobacco	Entertain.	Travel	Donations	Eating Out	HH total food
No food 0-5	-14.513***	-6.079**	-1.183	-43.262	8.283	-6.289***	-27.088***
	[-3.681]	[-2.549]	[-0.231]	[-1.156]	[0.888]	[-3.819]	[-3.807]
No food 6-12	5.067	1.266	3.455	-15.062	-6.277	-0.382	10.153
	[1.320]	[0.576]	[0.702]	[-0.352]	[-0.646]	[-0.212]	[1.490]
No food 13-17	1.589	3.020	0.500	-53.321	15.655	-1.229	4.906
	[0.288]	[1.106]	[0.095]	[-1.452]	[1.080]	[-0.825]	[0.547]
Marital status	9.098	3.219	1.707	39.287	9.880	-3.003*	16.844
	[1.007]	[0.869]	[0.190]	[0.271]	[1.111]	[-1.749]	[1.242]
Age	-1.229***	-1.243***	0.453**	2.278	-0.027	-0.867***	-3.978***
	[-4.548]	[-9.290]	[2.004]	[0.793]	[-0.046]	[-8.873]	[-8.668]
Gender	-1.000	5.083***	-3.411	-98.598**	-2.482	1.404	6.965
	[-0.315]	[2.812]	[-0.794]	[-2.431]	[-0.173]	[0.823]	[1.183]
Rural	-101.989***	-2.680	-19.993***	-707.642***	-22.325	-18.821***	-73.048***
	[-23.239]	[-0.880]	[-5.893]	[-10.918]	[-1.551]	[-5.875]	[-8.904]
Educ1	-102.288***	6.092	9.192	-3,449.420***	-39.864	-70.299***	-176.233***
	[-5.401]	[0.822]	[0.302]	[-4.515]	[-0.505]	[-4.468]	[-4.844]
Educ2	-57.445***	1.673	2.567	-2,963.761***	-33.026	-46.182***	-121.578***
	[-2.932]	[0.211]	[0.082]	[-3.829]	[-0.401]	[-2.822]	[-3.281]
HH Total Income	0.274***	0.131***	0.068***	4.953***	1.610*	0.284***	0.680***
	[4.758]	[3.362]	[3.684]	[3.993]	[1.847]	[3.767]	[4.704]
HH size	24.333***	4.282***	0.577	-11.994	-6.198	0.805	36.109***
	[19.114]	[6.653]	[0.360]	[-0.913]	[-0.877]	[1.209]	[16.056]
HH financial assets	0.387***	-0.010	0.009**	13.287***	0.076	0.009	0.694*
	[3.574]	[-1.193]	[2.517]	[57.658]	[0.127]	[0.502]	[1.801]
Year FE	Y	Y	Y	Y	Y	Y	Y
Observations	29,257	29,208	29,339	29,208	29,420	29,339	29,208
R-squared	0.068	0.016	0.005	0.113	0.018	0.045	0.065
AIC	415,502	393,072	438,914	588,336	502,835	390,463	429,836
BIC	415,610	393,180	439,022	588,444	502,943	390,571	429,943

Independent variable *No food 0-5*, *No food 6-12, and No food 13-17* captures the differential impact of famine based on the individual's age at which they experienced food shortage.

As we can see in Panel A, the estimated coefficients on *No food 0-5* is negative and statistically significant at 1% level in all food related expenditures (Columns 1, 6 and 7), and magnitude varies from -6.289 (eating out expenditures) to -27.088 (household total food expenditures). Among other expenditures, it is only alcohol and tobacco expenditures which are negatively correlated and show statistical significance (at 5%). Entertainment and travel expenditures (Columns 3 and 4) are negatively correlated, yet do not show statistical significance at conventional levels, while charitable donations expenditures are positively correlated yet imprecisely estimated.

Estimated coefficients on two other age groups *No food 6-12, and No food 13-17* provide mixed evidence, when it comes to food expenditures (Columns 1, 6 and 7). Household food expenditures irrespective of whether they include or not eating out expenditures are positive (Columns 1 and 7), while eating out expenditure has a negative sign. Having said that, none of the coefficient is significant at conventional levels.

This suggests that those respondents who were between the ages of 0 and 5 at the time of exposure to food shortage tend to have the lowest Food Expenditures out of all age groups, suggesting that the memory of famine has the largest effect on this age group. This is in line with our findings from previous chapter.

In Panel B (*Table 3.16*), when I include both year and province fixed effects, I find similar results.

Table 3. 16: Effects of not having enough food in different age periods on household expenditures; Year and Province FE

Explanatory Variables are defined as described in Table 3.1; No food 0-5 presents whether respondent's family while he/she was between 0 and 5 did not have enough to eat (1) or that has not happened (0); No food 6-12 presents whether respondent's family while he/she was between 6 and 12 did not have enough to eat (1) or that has not happened (0); No food 13-17 presents whether respondent's family while he/she was between 13 and 17 did not have enough to eat (1) or that has not happened (0);
Dependent variables are defined as follows Food excl. eating out represents household food expenditures in the past 7 days which exclude eating out expenditures (yuan); Alcohol and tobacco represents household expenditures in the past 7 days on alcohol and tobacco (yuan); Eating out represents household expenditures in the past 7 days on food eaten away from home (yuan); HH total food represents household expenditures in the past 7 days on purchased food and food eaten from own production, meals eaten out as well as on alcohol and tobacco (yuan); Entertain., Travel and Donations represents expenditures of a household in the past 30 days for entertainment purposes, non-work related travel and charitable donations.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include year and province fixed effects.*

Panel B	(1) Food excl. eating out	(2) Alcohol and tobacco	(3) Entertain.	(4) Travel	(5) Donations	(6) Eating Out	(7) HH total food
No food 0-5	-14.100***	-6.626***	-2.255	-54.467	7.947	-6.340***	-28.704***
	[-3.355]	[-2.669]	[-0.432]	[-1.463]	[0.815]	[-3.703]	[-3.807]
No food 6-12	6.107	1.274	4.160	-5.365	-6.289	-0.206	10.272
	[1.597]	[0.582]	[0.825]	[-0.122]	[-0.664]	[-0.120]	[1.521]
No food 13-17	0.340	1.850	-0.386	-53.540	11.826	-1.462	-0.426
	[0.065]	[0.702]	[-0.073]	[-1.440]	[0.860]	[-0.962]	[-0.050]
Marital status	9.845	3.200	5.351	44.760	18.194**	-3.380*	18.218
	[1.094]	[0.863]	[0.599]	[0.318]	[2.211]	[-1.960]	[1.345]
Age	-1.408***	-1.230***	0.622**	1.117	0.217	-0.846***	-4.027***
	[-5.131]	[-9.389]	[2.572]	[0.437]	[0.442]	[-8.416]	[-8.694]
Gender	-0.708	4.849***	-3.707	-93.889**	-3.895	1.346	6.422
	[-0.224]	[2.678]	[-0.865]	[-2.277]	[-0.272]	[0.790]	[1.095]
Rural	-101.179***	-5.079	-18.000***	-663.475***	-32.557**	-17.811***	-72.863***
	[-22.619]	[-1.592]	[-5.033]	[-10.827]	[-2.570]	[-5.664]	[-8.726]
Educ1	-100.443***	3.682	11.610	-3,270.788***	-53.279	-69.740***	-177.499***
	[-5.453]	[0.499]	[0.374]	[-4.315]	[-0.718]	[-4.421]	[-4.899]
Educ2	-54.141***	1.266	2.972	-2,764.645***	-40.778	-45.552***	-116.594***
	[-2.835]	[0.161]	[0.094]	[-3.604]	[-0.502]	[-2.779]	[-3.158]
HH Total Income	0.251***	0.126***	0.082***	4.695***	1.631*	0.279***	0.650***
	[4.527]	[3.232]	[3.594]	[3.774]	[1.829]	[3.682]	[4.569]
HH size	24.161***	4.338***	-0.253	-19.910	-9.653	1.067	35.693***
	[18.362]	[6.513]	[-0.158]	[-1.502]	[-1.109]	[1.453]	[15.316]
HH financial assets	0.353***	-0.012	0.008*	13.174***	0.104	0.006	0.667*
	[3.263]	[-1.491]	[1.741]	[64.084]	[0.164]	[0.382]	[1.714]
Year FE	Y	Y	Y	Y	Y	Y	Y
Province FE	Y	Y	Y	Y	Y	Y	Y
Observations	29,257	29,208	29,339	29,208	29,420	29,339	29,208
R-squared	0.078	0.021	0.014	0.119	0.020	0.048	0.074
AIC	415,230	392,979	438,693	588,159	502,824	390,432	429,617
BIC	415,562	393,311	439,026	588,491	503,157	390,764	429,947

4.5.4 Alternative proxy for famine exposure: starvation to death during the Great Chinese Famine

To establish whether the effect of prior exposure to famine on individual's lifestyle choices later in life reported above is sensitive to measurement issues, I also use a third measure for famine exposure: $Starved\ to\ Death$. The main independent variables of interest $Starved\ to\ Death$ equals 1 if there was someone in respondent's household who starved to death during the Great Chinese Famine 1958-62, and 0 otherwise:

$$Starved\ to\ Death = \begin{bmatrix} 1 & if\ in\ respondent's\ household\ someone\ starved\ to\ death\ during\ Great\ Famine \\ 0 & otherwise \end{bmatrix}$$

Using this alternative measure of exposure to famine, I re-estimate Equation (1). Results are reported in Table *Table 3.17*.

As before, Panel A shows the results that include year fixed effects, and Panel B (*Table 3.18*) shows the results that include both year and province fixed effects.

As we can see from Panel A, the estimated coefficient on *Starved to Death* is positive and statistically significant at 5% level in case on Entertainment spending (Column 3). It is also positive and statistically significant at 10% level in case of charitable Donations (Column 5), suggesting that those respondents whose have household members who starved to death during the Great Chinese Famine, tend to spend more on entertainment and charitable donations later in life. At the same time, as we can see from Column 6, they tend to spend less on Eating Out Expenditures, as indicated by a negative and statistically significant coefficient in Column 6. Similar results are obtained in Panel B (*Table 3.18*), with the inclusion of both year and province fixed effects.

Table 3. 17: Effects of family member starving to death during the Great Famine on household expenditures; Year FE

Explanatory Variables are defined as described in Table 3.1; Starved to death presents whether any family member starved to death during the Great Famine (1) or not (0).

Dependent variables are defined as follows Food excl. eating out represents household food expenditures in the past 7 days which exclude eating out expenditures (yuan); Alcohol and tobacco represents household expenditures in the past 7 days on alcohol and tobacco (yuan); Eating out represents household expenditures in the past 7 days on food eaten away from home (yuan); HH total food represents household expenditures in the past 7 days on purchased food and food eaten from own production, meals eaten out as well as on alcohol and tobacco (yuan); Entertain., Travel and Donations represents expenditures of a household in the past 30 days for entertainment purposes, non-work related travel and charitable donations.

*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include year fixed effects.*

Panel A	(1) Food excl. eating out	(2) Alcohol and tobacco	(3) Entertain.	(4) Travel	(5) Donations	(6) Eating Out	(7) HH total food
Starved to death	-4.141	-1.299	8.044**	-32.439	29.059*	-3.998**	-3.988
	[-0.822]	[-0.506]	[2.067]	[-0.747]	[1.774]	[-2.471]	[-0.501]
Marital status	9.264	3.280	1.787	38.140	9.511	-3.034*	17.115
	[1.018]	[0.881]	[0.199]	[0.261]	[1.093]	[-1.774]	[1.254]
Age	-1.173***	-1.197***	0.453**	1.742	0.048	-0.868***	-3.860***
	[-5.051]	[-9.898]	[2.119]	[0.595]	[0.096]	[-9.042]	[-9.493]
Gender	-1.222	5.009***	-3.252	-102.034**	-2.065	1.138	6.654
	[-0.384]	[2.737]	[-0.753]	[-2.467]	[-0.149]	[0.665]	[1.121]
Rural	-102.393***	-2.762	-20.018***	-713.945***	-21.731	-19.233***	-73.738***
	[-23.208]	[-0.909]	[-5.895]	[-11.012]	[-1.428]	[-6.029]	[-8.960]
Educ1	-101.984***	6.204	9.604	-3,458.601***	-40.228	-70.784***	-175.585***
	[-5.373]	[0.835]	[0.317]	[-4.519]	[-0.506]	[-4.497]	[-4.822]
Educ2	-57.890***	1.392	2.908	-2,968.488***	-32.908	-46.676***	-122.411***
	[-2.946]	[0.175]	[0.094]	[-3.830]	[-0.397]	[-2.855]	[-3.302]
HH Total Income	0.275***	0.131***	0.068***	4.963***	1.610*	0.285***	0.681***
	[4.759]	[3.361]	[3.691]	[4.003]	[1.849]	[3.768]	[4.700]
HH size	24.333***	4.287***	0.591	-12.562	-6.172	0.774	36.128***
	[19.104]	[6.649]	[0.368]	[-0.965]	[-0.880]	[1.169]	[16.061]
HH financial assets	0.392***	-0.010	0.009**	13.291***	0.073	0.009	0.704*
	[3.615]	[-1.180]	[2.469]	[57.429]	[0.124]	[0.519]	[1.824]
Year FE	Y	Y	Y	Y	Y	Y	Y
Observations	29,257	29,208	29,339	29,208	29,420	29,339	29,208
R-squared	0.067	0.016	0.005	0.112	0.018	0.045	0.064
AIC	415,511	393,075	438,909	588,334	502,830	390,470	429,845
BIC	415,602	393,167	439,000	588,426	502,922	390,562	429,936

Table 3. 18: Effects of family member starving to death during the Great Famine on household expenditures; Year and Province FE

Explanatory Variables are defined as described in Table 3.1; Starved to death presents whether any family member starved to death during the Great Famine (1) or not (0).

Dependent variables are defined as follows Food excl. eating out represents household food expenditures in the past 7 days which exclude eating out expenditures (yuan); Alcohol and tobacco represents household expenditures in the past 7 days on alcohol and tobacco (yuan); Eating out represents household expenditures in the past 7 days on food eaten away from home (yuan); HH total food represents household expenditures in the past 7 days on purchased food and food eaten from own production, meals eaten out as well as on alcohol and tobacco (yuan); Entertain., Travel and Donations represents expenditures of a household in the past 30 days for entertainment purposes, non-work related travel and charitable donations.

*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include year and province fixed effects.*

Panel B	(1) Food excl. eating out	(2) Alcohol and tobacco	(3) Entertain.	(4) Travel	(5) Donations	(6) Eating Out	(7) HH total food
Starved to death	-4.699	-2.342	9.759**	2.582	28.565	-4.575***	-8.519
	[-0.875]	[-0.822]	[2.412]	[0.058]	[1.506]	[-2.712]	[-0.951]
Marital status	10.001	3.231	5.445	43.088	17.914**	-3.447**	18.327
	[1.105]	[0.867]	[0.610]	[0.303]	[2.214]	[-2.006]	[1.345]
Age	-1.359***	-1.194***	0.615***	0.565	0.256	-0.848***	-3.953***
	[-5.704]	[-10.043]	[2.663]	[0.211]	[0.600]	[-8.455]	[-9.564]
Gender	-0.902	4.723**	-3.554	-96.963**	-3.562	1.081	5.917
	[-0.285]	[2.575]	[-0.824]	[-2.303]	[-0.258]	[0.634]	[1.000]
Rural	-101.572***	-5.258*	-18.029***	-669.215***	-32.058**	-18.194***	-73.837***
	[-22.630]	[-1.659]	[-5.074]	[-10.913]	[-2.435]	[-5.812]	[-8.847]
Educ1	-99.991***	3.697	12.058	-3,278.456***	-53.826	-70.161***	-177.152***
	[-5.419]	[0.501]	[0.390]	[-4.319]	[-0.724]	[-4.449]	[-4.887]
Educ2	-54.490***	0.938	3.274	-2,768.837***	-40.775	-46.025***	-117.618***
	[-2.847]	[0.119]	[0.104]	[-3.606]	[-0.499]	[-2.811]	[-3.185]
HH Total Income	0.252***	0.126***	0.082***	4.705***	1.630*	0.280***	0.652***
	[4.532]	[3.233]	[3.595]	[3.782]	[1.830]	[3.683]	[4.568]
HH size	24.190***	4.349***	-0.233	-20.021	-9.650	1.063	35.762***
	[18.381]	[6.524]	[-0.145]	[-1.512]	[-1.110]	[1.447]	[15.327]
HH financial assets	0.358***	-0.012	0.008*	13.179***	0.100	0.007	0.680*
	[3.308]	[-1.467]	[1.703]	[63.911]	[0.159]	[0.401]	[1.743]
Year FE	Y	Y	Y	Y	Y	Y	Y
Province FE	Y	Y	Y	Y	Y	Y	Y
Observations	29,257	29,208	29,339	29,208	29,420	29,339	29,208
R-squared	0.078	0.021	0.014	0.119	0.020	0.047	0.073
AIC	415,238	392,983	438,688	588,158	502,820	390,439	429,628
BIC	415,553	393,299	439,004	588,474	503,136	390,755	429,941

I next proceed to re-estimate Equation (2), to analyse the effect of prior exposure to famine on family planning decisions using *Starved to Death* as an alternative exposure measure. Results are reported in *Table 3.19*. As before, Panel A shows the results that include year fixed effects, and Panel B shows the results that include both year and province fixed effects.

Table 3. 19: Effects of family member starving to death during the Great Famine on family planning decisions; linear probability model; Year FE

Explanatory Variables are defined as described in Table 3.1; Starved to death presents whether any family member starved to death during the Great Famine (1) or not (0).
Dependent variables are defined as described in Table 3.4
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include year fixed effects.*

	(1)	(2)	(3)	(4)
Panel A	Put off marriage	Put off birth	Couldn't give birth	Artificial abortion
Starved to death	0.019***	0.027***	0.026***	0.018***
	[5.082]	[6.737]	[7.262]	[6.349]
Marital status	-0.010***	-0.008**	-0.002	-0.008***
	[-3.055]	[-2.457]	[-0.822]	[-3.804]
Age	0.000***	0.001***	0.001***	0.000**
	[4.155]	[6.061]	[7.688]	[2.361]
Gender	0.009***	0.004**	0.002	-0.001
	[4.348]	[2.269]	[1.482]	[-0.685]
Rural	0.011***	0.008***	0.008***	0.006***
	[5.085]	[3.817]	[4.803]	[5.525]
Educ1	-0.021*	-0.012	0.003	-0.016**
	[-1.853]	[-1.191]	[0.543]	[-2.046]
Educ2	-0.022*	-0.016*	-0.003	-0.016**
	[-1.924]	[-1.651]	[-0.562]	[-2.084]
HH Total Income	0.000	-0.000*	0.000	-0.000
	[0.012]	[-1.940]	[0.099]	[-1.490]
HH size	0.002***	0.001	0.001**	-0.000
	[2.844]	[1.220]	[2.362]	[-0.239]
HH financial assets	-0.000	-0.000**	-0.000	-0.000*
	[-0.960]	[-2.004]	[-0.590]	[-1.716]
Year FE	Y	Y	Y	Y
Observations	30,270	30,270	30,270	30,270
R-squared	0.005	0.006	0.008	0.006
AIC	-23,256	-26,028	-35,220	-55,978
BIC	-23,164	-25,937	-35,128	-55,887

The main independent variable of interest *Starved to Death is* defined as before. As we can see from Column 1 in Panel A, where the dependent variable is *Put off Marriage*, the estimated coefficient on

Starved to Death is positive and statistically significant at 1% level, suggesting that respondents' whose household member experienced death due to starvation during the Great Chinese Famine tend to be more likely to delay getting married, controlling for a host of individual and household characteristics. Dependent variable in Column 2 is *Put off baby*: the estimated coefficient on *Starved to Death* is again positive, and also statistically significant at 1% level, suggesting that prior exposure to famine increases the probability of delaying having children later in life. Column 3 reports the results for inability to have children as the dependent variable - the estimated coefficient on *Starved to Death* is again positive, and also statistically significant at 1% level, suggesting that individuals whose household members experienced death due to starvation during the Great Chinese Famine tend to be less likely to be able to have children later in life. In Column 4 the dependent variable is *Artificial Abortion* - the estimated coefficient on *Starved to Death* is again positive, and statistically significant at 1% level, suggesting that prior exposure to famine increases the probability of having an abortion later in life.

In Panel B (*Table 3.20*), I repeat the analysis, but now including both year and province fixed effects. This ensures that I control for any time- and province-specific unobservable factors that can be driving my results. The results remain unchanged, both in terms of the economic magnitudes and statistical significance. As it can be seen from Panel B, respondents whose families experienced death due to starvation during the Great Famine, tend to be more likely to delay getting married and having children, as indicated by positive and highly significant coefficient on *Starved to Death* (Columns 1 and 2). They are also less likely to have children, and more likely to have an abortion (Columns 3 and 4).

Table 3. 20 Effects of family member starving to death during the Great Famine on family planning decisions; linear probability model; Year and Province FE

Explanatory Variables are defined as described in Table 3.1; Starved to death presents whether any family member starved to death during the Great Famine (1) or not (0).
Dependent variables are defined as described in Table 3.4
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include year and province fixed effects.*

	(1)	(2)	(3)	(4)
Panel B	Put off marriage	Put off birth	Couldn't give birth	Artificial abortion
Starved to death	0.018***	0.024***	0.024***	0.018***
	[4.626]	[5.819]	[6.338]	[6.113]
Marital status	-0.007**	-0.005*	-0.000	-0.008***
	[-2.248]	[-1.679]	[-0.177]	[-3.496]
Age	0.000***	0.001***	0.001***	0.000**
	[4.131]	[6.041]	[8.778]	[2.398]
Gender	0.009***	0.004**	0.002	-0.001
	[4.303]	[2.208]	[1.349]	[-0.779]
Rural	0.008***	0.005**	0.006***	0.006***
	[3.364]	[2.137]	[3.666]	[5.476]
Educ1	-0.022**	-0.013	0.003	-0.016**
	[-1.972]	[-1.365]	[0.565]	[-2.079]
Educ2	-0.022**	-0.016	-0.002	-0.016**
	[-1.962]	[-1.634]	[-0.294]	[-2.018]
HH Total Income	0.000	-0.000	0.000	-0.000
	[0.155]	[-1.429]	[0.588]	[-1.462]
HH size	0.001	-0.000	0.001	-0.000
	[1.362]	[-0.557]	[1.211]	[-0.894]
HH financial assets	-0.000	-0.000*	0.000	-0.000
	[-1.041]	[-1.763]	[1.212]	[-1.243]
Year FE	Y	Y	Y	Y
Province FE	Y	Y	Y	Y
Observations	30,270	30,270	30,270	30,270
R-squared	0.012	0.012	0.018	0.010
AIC	-23,435	-26,140	-35,461	-56,057
BIC	-23,119	-25,824	-35,144	-55,741

Based on information criteria tests (AIC) specifications which include both province and wave fixed effect seem to perform better relative to those that include only wave fixed effect. Based on information criteria tests (BIC) specifications which include only wave fixed effect seem to exhibit better relative goodness-of-fit comparing to those that include both province and wave fixed effect. This holds irrespective of how I measure famine exposure.

5 Discussion

In the previous three sections I investigate relationship between exposure to famine and nutrient intake from several different angles. China Health and Nutrition Survey (CHNS), which I use in *Section 2* and *Section 3* does not contain explicit information on exposure to famine, and I use proxy measure. As the proxy for the exposure to famine in those two sections, I use Excess Death Rate (EDR), calculated in two different ways. This process is elaborated in the *Methodology* sub-section of *Section 2*. In *Section 4,* where I use China Health and Retirement Longitudinal Study (CHARLS) I use three metrics for exposure to famine and food shortage. However, in this case all three metrics are derived directly from respondents' answers, hence no proxy measure was used. The three questions, which were used for constructing the three metrics provide information on the exposure to food insecurity and hunger, and each is associated to a different degree of hunger severity. The question about the occurrence of food shortage episodes before age 17 reflects first-degree hunger, which is the mildest form. Additionally, the question about starvation during the Great Famine reflects second-degree famine. Finally, the question about death of a family member caused by starvation during the Great Famine indicates third-degree famine, the most severe form. Therefore, the three measures of food shortage are "Not enough food before age 17", "Starvation during the Great Famine", "Death caused by the Great Famine".

Additionally, nutrient intake was also measured in different ways. In *Section 2*, nutrient intake is proxied by Dietary Diversity Score (DDS) that is calculated in three different ways, using six, twelve, and twenty food groups. In *Section 3*, nutrient intake is proxied by absolute macronutrient intake, as well as relative share of macronutrients in total diet, where macronutrients are fat, protein and carbohydrates. In *Section 4*, I proxy nutrient intake by share of total household food expenditures as well as eating out expenditures.

Finally, when it comes to CHNS dataset, in the first iteration, I used only four waves of CHNS (2004, 2006, 2009, 2011) both as a cross-section and a pooled cross-sectional analysis. In the following iteration, I added two more waves (1997, 2000) to the pooled cross-sectional analysis. Temporal dimension in CHARLS dataset is uniform, as it is a longitudinal dataset.

I apply different metrics in the dissertation is to test the sensitivity of models to choice of dependent and independent variables as well as choice of period captured by the analysis. The results from the three sections are discussed below.

The results from *Section Two* show that irrespective of number of food groups used in the model, the unit for measuring hunger (EDR [%] or ERD [diff]) or the application of cross-sectional or pooled analysis, **the relationship between exposure to famine and dietary diversity score** (DDS) is consistent. It **is negative** and significant at the 1% level. Adding two more waves (1997 and 2000) to the pooled regression did not change the results.

The results from *Section Three* suggest that **exposure to famine is negatively correlated with quantity** of calories, fat, protein and carbohydrates consumed during the day (at 1%). When it comes to the quantitative composition of diet, the results are more complex. **Higher exposure to famine** is **positively** correlated with **share of carbohydrates** in diet (at 1%) and **negatively** correlated with **share of protein** (at 1%) and **share of fat** (at 10%). In addition to this, when it comes to DDS, the results suggest that a higher share of carbohydrates in diet has a negative effect on DDS (at 1%), while a higher share of fat and protein in diet have positive effects (at 1%).

In *Section Four*, where I use the three metrics of food shortage, the results show that all three measures have negative effect on nutrient intake, measured by **household food expenditures**. The results hold irrespective of whether those expenditures include eating away from home or not. Having said that, the significance level varies. The coefficient is significant at 1% when food shortage is measured by "Starvation during the Great Famine", while the coefficients of the two other metrics are imprecisely estimated. Interestingly, estimated coefficients for regional dummy variables suggest that there are certain unobservable province characteristics which negatively affect household food expenditures in provinces which were more affected by the Great Famine. Results from the analysis of impact of food shortage on "**eating out expenditures**" show the opposite pattern from the one described above. This time, estimations that include "Not enough food before Age 17" as well as "Death caused by the Great Famine" are negative and significant at 1%, while "Starvation during the Great Famine" is not significant at conventional levels. Taken together, **exposure to famine** on average **negatively affects food expenditures** of those who were exposed to famine either directly or through parents, siblings or grandparents. However, note that different measures of exposure to famine provide different evidence and using only one measure is insufficient to understand the complexity of the dynamic relationship between famine and nutrient intake, measured by food expenditures. Those results confirm the *Chapter One* and *Chapter Two* findings.

Taken together, the results from Section 2, 3, and 4 suggest that comparing to those who are not affected by food shortage, those who are affected by food shortage, spend less money on food, their food basket contain less diverse foods. Also, they consume diets with higher share of carbohydrates and lower share of fat and protein. In other words, populations affected by food shortage consume poorer diets. From the nutrition transition perspective, the are in earlier phase of nutrition transition relative to those not affected by food shortage. These findings cannot be related to the current literature, as to the best of my knowledge, this is the first research which analyses long-term famine effects on nutrient intake.

When it comes to other explanatory variables the literature that explores the relationship between age and DDS offers mixed evidence. My estimations in *Section Two* show that the linear relation between age and dietary diversity score has a positive coefficient, significant at 5% level in 2006 and negative and significant at 5% and 10% levels in 2011, irrespective of the EDR measure. While in a pooled regression (four waves), the coefficient is positive yet not significant at conventional levels, in pooled regression (six waves), the coefficient is positive and significant at 10%. This led me to explore the quadratic relationship between age and EDR, and the results suggest that there is a negative and significant relationship between the two only in 2004, irrespective of EDR metrics. A pooled regression (four waves) not only confirmed this, but the significance level increased from significant at 10% in 2004 to significant at 1%. Repeating the same analysis using pooled regressions (six waves) revealed that the relationship is negative, irrespective of number of food groups used as DDS proxy. However, this relationship is significant at 1% when I use 20 food groups, significant at 5% when I use 12 food groups and statistically insignificant when I use six food groups as a measure of DDS. This is the evidence that the **relationship between age and DDS is not linear but rather convex** and we also see that choice of metrics used might drive significance level and magnitude of coefficients. In the following stage, I investigate effects of famine on specific age cohorts, and I found the exact birth-year affected most by the famine, measured by DDS later in life. Cross-sectional regressions using both EDR [%] and EDR [diff] are consistent and combined with pooled regression (1997–2011) strongly suggest that **those born in 1957 are mostly affected by the famine**, measured by DDS. Since the famine reached different provinces at different times, some in 1959 and some later, the current dataset does not allow precise estimation how old the famine survivor was when famine was the most severe, and that person could be between two and five years old.

Results from the *Section Three* indicate similar results. There, I also explore heterogenous effects of exposure to famine on the quantitative and qualitative composition of diets, by birth-year cohorts. The results suggest that, compared to those born after famine, *Cohort 1956* has lower protein intake (at 10%) and caloric intake (at 5%); *Cohort 1959* has lower carbohydrate intake (at 10%), *Cohort 1960* has lower protein, carbohydrate and caloric intake (at 10%, 1% and 10% respectively) and Cohort 1962 has lower protein, carbohydrate and caloric intake (at 5%, 1% and 10% respectively).

Results related to qualitative diet composition suggest that only Cohort 1962 has lower share of fat in diet (at 5%).

Similarly, my analysis of effect of food shortage on different age cohorts In *Section Four* shows that those who experienced food **shortage** when they were **5 years old or younger are most affected**. Both food and non-food expenditures are negatively correlated with this age group, except for donations expenditures (not statistically significant). As for the food expenditures, irrespective of whether those expenditures include eating away from home, they are negative and significant at 1%.

Taken together, the results from Section 2, 3, and 4 suggest that exposure to famine does not affect all age groups in the same way, and this confirms findings from the literature (X. F. Hu et al., 2017; Y. Li et al., 2010; Ning et al., 2019; Qin et al., 2020; Sparen et al., 2004). Furthermore, I find that those who are 5 years old or younger during the famine have spend less money on food and have poorer diet later in life than those born after famine. My results support the findings of Qin et al. (2020), who found that exposure to famine during the foetal development and childhood has a higher risk of metabolic syndrome and contradicts results from Ning et al. (2019). The similarities and differences between my results, and results from other studies should be taken with caution, as I study effects of famine on nutrient intake, and study effects of famine on MetS.

In addition to this, **Interaction** between **EDR** and **Age squared** has a **positive** coefficient both in the cross-sectional and pooled analysis of *Section Two*. Having said that the statistical significance is higher in pooled regressions (1%) then in cross-sectional ones (at 10% or no statistical significance). Comparing EDR metrices, EDR [%] consistently show higher magnitude and lower statistical significance than EDR [diff]. Here, we see that choice of metrics might play a role in statistical significance of the results.

With respect to income, the results from *Section Two* suggest that, irrespective of the measure of EDR, the wave, cross-sectional or pooled regressions, **the relationship between income and DDS is consistent: it is positive** and significant at 1% level. This is very much in line with the literature.

In Section Three I find that ***household Income*** is positively correlated with total intake of calories as well as intake of fat and protein (all at 1%). While income is negatively associated with intake of carbohydrates, it is not significant at conventional levels. Additionally, higher household income is positively correlated with share of fat and protein in diet (both at 1%) and negatively correlated with intake of carbohydrates (at 1%).

In the same vein *Section Four* identifies that **Household income** is positively associated with food expenditures in general (at 1%).

In addition to this, in *Section Two* I also investigate **interaction** between **inverted EDR** and **Income,** and its relationship with DDS. The results show that there is a **positive** relationship between the two. In the cross-sectional analysis, this coefficient is significant at 5% only in the 2004 wave. As for the pooled regressions, the coefficient shows higher statistical significance in 1997–2011 (at 1%) than in 2004–2011 (5%) and when using EDR [diff] metrics (at 1%), compared to EDR [%] metrics (at 5%). Finally, comparing the famine cohort to the post-famine cohort, this interaction shows a diminishing effect. These findings suggest that higher income mitigates the negative effects of famine.

Taken together, evidence from *Sections Two, Three and Four* are consistent when it comes to impact of income on nutrient intake. Irrespective of whether one was exposed to food shortage or not, this relationship is positive. These results confirm findings from the current literature (Babalola & Isitor, 2014; Doan, 2014; Hou et al., 2021; Wang et al., 2021; Zani et al., 2019; Zhang et al., 2017; Zhao et al., 2020)

Section Two found that the **relationship between education and DDS** is also **positive**. Yet different approaches to estimations provided slightly different results. While the sign and magnitude are similar between EDR [%] and EDR [diff], EDR [%] ones show stronger confidence (at 1% level) compared to EDR [diff] one, which is significant at 5% level. On the other hand, pooled regressions show that the relationship is positive and significant at 1% level, irrespective of number of waves and EDR metrics.

As per *Section Three***, Education** is inversely related to intake of carbohydrates (at 1%) and directly correlated with fat intake (at 1%). Furthermore, more years of education assumes a lower share of carbohydrates and a higher share of protein and fat in diet (at 1%, 10% and 1% respectively).

Section Four results suggest that more educated respondents spend more money on food, relative to less educated. Coupled with evidence from the literature which argues that food expenditures are directly proportional to DDS, that indicates that more educated sub populations consume better quality diet.

In Section Two, I test **Interaction** between **inverted EDR** and **Education, and I find that it** has a **positive** coefficient and follows the same pattern as the interaction with *Income* described above. The coefficient has higher statistical significance in the case of a pooled regression analysis, compared to a cross-sectional one, and in the case of EDR [diff], compared to EDR [%], and a diminishing effect in post-famine generations.

Similarly to impact of income on nutrient intake, results from *Sections Two, Three and Four* are consistent when it comes to impact of education on nutrient intake. All three sections find a positive relationship between the two. From the perspective of the existing literature, while some authors also find positive relationship between the two (Bi et al., 2019; Hou et al., 2021; Zhong et al., 2018), others find the negative relationship (Sotsha et al., 2019; Zani et al., 2019). While Sotsha et al. (2019) do not comment on this, Zani et al. (2019) postulate that more educated individuals ,might show more efficient food shopping habits, which in turn led to decrease in food expenditures.

When it comes to the effects of household size on nutrient intake, the results are consistent across *Sections 2-4*. Results from *Section Two* show that **Household size has a negative impact on DDS**, and while it is consistent in 2004 and 2006, this effect diminishes in the 2009 and 2011 waves. In pooled regressions, including 2004–2011 and 1997–2011, irrespective of number of food groups used to measure DDS, the effect is negative and significant at 1%.

As per Section Three**, Household size** is positively associated to total intake of carbohydrates and protein but negatively associated to total fat intake (at 1%, 10% and 1%). However, share of carbohydrates is positively correlated with household size (at 1%), while fat share is negatively correlated (at 1%).

Results from Section Four suggest that Household size is positively associated to total food expenditures and total food expenditures which exclude away from home food expenditures (at 1%).

Taken together, these results show that larger households spend more money on food, but the quality of those foods are lower comparing to smaller households.

Results of Section Two suggest that the **relationship between marital status and DDS is a positive one**, and it is very similar to household size, yet the sign is the opposite. Namely, the relationship is positive and significant at 1%, 5% and 10% in 2004 and 2006, where the significance level depends on EDR metrics. This effect diminishes in the 2009 and 2011 waves.

The results from *Section Three* confirm those results. Namely, **Married** people on average have a lower share of carbohydrates and a higher share of protein and fat in their diet, relative to the unmarried (at 1%). In other words, they have higher quality diet relative to those who are not married.

Another explanatory variable which I use in my analysis as a control variable is urban/rural residence a respondent. In Section Two, the coefficient is positive, irrespective of EDR metrics, DDS measure or numbers of waves used. However, while the coefficient is positive and significant at 1% in case of pooled regressions, in case of cross sectional is significant at different levels only in some waves.

Section Three suggest that **Urban** residents consume fewer calories and carbohydrates (at 1%), compared to rural residents. As for the qualitative characteristics of diet, share of carbohydrates is lower and share of protein higher in diets of urban residents than of those who live in rural areas.

The results from Section Four comport with the previous two Sections and show that living in rural areas is negatively associated to food expenditures in general (at 1%). Those results, taken together with the results from Section Two and Section Three, imply that urban residents consume better quality diet, which confirms findings from the literature (Doan, 2014; Wang et al., 2021; Zhang et al., 2017; A. Zhao et al., 2020).

Taken all discussed results together, it becomes clear that findings are consistent when it comes to evaluating the effects of food shortage on nutrient intake. Furthermore, other control variables such as income, household size, education and urban residence show high degree of consistency. Having said that, the results also revealed that choice of metrics can influence statistical significance and magnitude of the coefficients.

Certain variables used in the analysis are common for *Sections Two, Three and Four*, and they are discussed above. Some other variables are specific for Sections Two and Four and will be presented below. Combined with the variables above, they will contribute to more comprehensive understanding of long-term effects of famine.

“Old home” refers to the father’s home, and the results from *Section Two* suggest that there is **no relationship between old home food habits and DDS**, neither in the cross-sectional nor pooled regressions. Having said that, only pooled regressions, which include a province dummy, indicate that there is a **positive correlation between old home food habits and DDS**. This, coupled with the fact that 96% of the sample stated that they keep old home food habits, suggests that these habits might play a significant role.

Belonging to the **Han nationality is negatively correlated with DDS**. While this relationship has not been found in 2004 and 2006, it has been identified in all other waves and in the cross-sectional and pooled regressions in *Section Two*. Having said that, including province dummies results in the statistical significance of the Han nationality disappearing. This result does not comport with the literature. This is likely driven by the fact that unlike the literature, I am being more cautious, as I restrict my sample only to non-movers, and this affects the relative share of Han nationals in the overall sample.

In the section below I present the results from *Section Four*.

In my analysis of **impact of famine on household non-food expenditures**, I also use all three measures of food shortage — “Not enough food before age 17”, “Starvation during the Great Famine”, “Death caused by the Great Famine starvation”. The analysed the expenditures are entertainment, nonwork-related travel, and charitable donations—in other words, non-essential expenditures.

This time, the results are more complex. Only in the case of **travel expenditures** are the results consistently negative, and only in the case of “Starvation during the Great Famine” they are statistically significant (at 5%), which indicates that individuals who starved during the Great famine spend less on non-work-related travel. This confirms findings from Yao et al. (2020).

In the case of **charitable donations**, only the “Not enough food before age 17” measure suggests there is a negative correlation between exposure to famine and charitable expenditures. However, the coefficient is not statistically significant. The other two measures suggest that those who starved during the famine and those who had a relative who died due to starvation show more sympathetic behaviour and donate more to charity. Between the two measures, results of those who experienced the third-degree famine are statistically significant at 5%. These results are to some extent in line with Y. Li et al. (2013) who find that after a natural disaster 9 year old children became more altruistic, and 6 year old children became less altruistic, yet they also find that 3 years after disaster, the altruistic behaviour returned to pre-disaster levels.

When it comes to **entertainment expenditures**, the results are not entirely in line with the literature. The relationship between exposure to famine and entertainment expenditures is negative, yet imprecisely estimated for “Not enough food before age 17” and “Starvation during the Great Famine”. When measuring exposure to famine by “Death caused by the Great Famine”, the results suggest that there is a positive and statistically significant relationship (at 5%), which indicates that those who lost parents, grandparents, siblings or children to the famine spend more on entertainment later in life. These findings certainly require deeper analysis.

Taken together, Section Four results suggest that **exposure to famine** on average **negatively affects travel expenditures** of those who were exposed to famine either directly or indirectly. For **other**

non-essential expenditures, the **findings are mixed** and are driven by the selection of a famine measure.

In my analysis of **impact of famine on family planning**, I also use all three measures of food shortage — "Not enough food before age 17", "Starvation during the Great Famine", "Death caused by the Great Famine". Family planning decisions include putting off marriage, putting of giving birth, inability to give birth and inducing abortion, all of them caused by food shortage in the Great Famine era.

As described earlier, the main rationale for exploring the relationship between exposure to famine and family planning decisions, is because the literature shows that Dietary Diversity Score of children ages six to 23 months is negatively associated with the mother's age (Dangura & Gebremedhin, 2017). In this case, all four family planning decisions might lead to having children at a later age, which translates to a lower DDS of the next generation.

All three have the same effect on decision to put off marriage. The coefficients are positive, significant at 1%, and have very similar magnitudes. This suggests that those who directly or indirectly **experienced food shortage** and famine are **more likely to put off marriage**, relative to those who were unaffected.

The following two **family planning dimensions**, to put off giving birth and inability to give birth, indicate similar pattern, based on famine measure. While they are all positive and significant at 1%, the magnitude differs across the types of famine measurement. First-degree hunger has the smallest magnitude, and third-degree hunger has the largest. In other words, when it comes to family planning characteristics, **the "deepest scar" is caused by the most severe exposure to famine**.

The decision to induce **artificial abortion** is positively associated to exposure to famine. All three metrics show a positive coefficient at 1%, where, "Starvation during the Great Famine" has the smallest magnitude, while "Death caused by the Great Famine starvation" has the largest. Hence, **the "deepest scar" is caused by the most severe exposure to famine**.

Overall, unlike household expenditures, the results related to the family planning decisions are uniform and consistent.

As elaborated earlier, I introduce province and wave fixed effects to control for time- and location-specific unobservable factors that can drive the relation of interest. In the case of measuring effects of famine measured by "Starvation during the Great Famine" on food expenditures, the magnitude and statistical significance decreased after introducing province fixed effects. The significance level decreased from 1% to 5% in case of food expenditures which both include and exclude food away from home.

In the case of the effects of famine measured by "Death caused by the Great Famine starvation" on food expenditures, the change occurred in the opposite direction. After introducing province fixed effects, the magnitude and statistical significance increased, with the significance changing from 5% to 1%.

Other control variables used in my **family planning** analysis suggest the following.

- Marital status is negatively correlated with family planning decisions (at 1%), suggesting that married respondents were less likely to put off giving birth or to pursue an artificial abortion than unmarried counterparts.
- Rural respondents were more likely to put off marriage, put off giving birth or not give birth and pursue an abortion (at 1%) than urban respondents.
- Household income did not play a role in family planning decisions, except in several regressions where it was negatively correlated with putting off birth (at 10%).
- Household size was positively correlated with putting off birth (at 1%) and with inability to give birth (at 1% and 5%).

Table 4.1 summarizes results obtained in Sections 2-4. The main purpose of the table to facilitate understanding of determinants of nutritional intake and in particular the effects of exposure to famine and food shortage episodes on diets. However, this table cannot capture all of the nuances presented in the results section of Sections 2-4. A comprehensive understanding of the effects of exposure to famine and food shortage episodes on diets requires a detailed analysis of the relationship between each determinant and nutritional outcome.

Table 4. 1: Summary of the results Chapter 2-4

	DDS	Diet quality	Food expenditure
Exposure to famine/ food shortage (EFFS)	-	-	-
Income	+	+	+
EFFS x Income	+		
Education	+	+	+
EFFS x Education	+	+	+
Age squared	-		
EFFS x Age squared	+		
Age 0-5	-	0	-
Household size	-	-	+
Married	+	+	-
'Old home' habits	+	0	
Urban residence	+	+	+
Share of carbohydrates	-		
Share of fat	+		
Share of protein	+		

"+" denotes positive relationship, *"-"* denotes negative relationship", *"0"* no relationship identified

6 Policy Implications

I explore the relationship between exposure to famine early in life and nutritional outcomes later in life. In other words, I explore the enduring effect of famine; the literature refers to this phenomenon as the “echo” or “scarring” effect.

For both famine and nutritional outcomes, I use several metrics as proxies. To measure famine, I use two Excess Death Rate (EDR) measures as well as answers-statements to three questions — whether someone did not have enough food to eat between age 0–17, whether one starved during the Great Famine and whether during the Great Famine one lost one or more family members due to starvation. To measure nutritional outcomes, I use Dietary Diversity Score (DDS), share of macronutrients in total diet, and food expenditures.

My results suggest that, irrespective of measure for famine and nutritional outcomes, exposure to famine early in life harms nutritional outcomes.

Nutritional outcomes play an important role in one’s health. As per literature reviewed in this dissertation, DDS is associated to nutrient and micronutrient adequacy, memory, cognitive function, depression, risk of being overweight, metabolic syndrome, cardio-metabolic risk factors, risk of fracture, a child’s nutritional status and growth and other health indicators (Li et al., 2021; Nguyen et al., 2018; Ruel, 2003; Wang et al., 2021; Yin et al., 2017; Zhang et al., 2017; Zhao et al., 2020; Zhao et al., 2017).

An extensive literature examines determinants of diet quality, and, to the best of my knowledge, this dissertation is the first attempt to examine the effect of exposure to famine on diet quality. My findings not only contribute to a better understanding of food consumption patterns but they may have several policy implications.

When it comes to health implications, my results suggest that those who were exposed to famine and food shortage earlier in life might have undetected predisposition to certain medical conditions associated with poor diet. Therefore, preventive medicine could treat famine survivors as a separate group. Regular medical examinations are critical in this context, as they could identify those medical conditions in early stages. This would contribute to reduction of undiagnosed and uncontrolled chronic conditions.

Furthermore, my results show that higher income mitigates the effects of famine. To that end, a poorer population exposed to food shortage earlier in life would benefit from some form of support in this respect, where food vouchers are one of the forms. The main purpose of those vouchers would not be to fulfil daily caloric requirements. Rather, their purpose would be to increase the dietary diversity of the target population. Similar programs, such as the well-established Women, Infants and Children (WIC) scheme, provide nutritious foods to the recipients. Some of the foods provided by WIC are dairy products, fruit, vegetables and eggs. Food items provided through similar programs would need to be adjusted to reflect habits and preferences of the population as well as to match food items that are locally available.

Education is another mitigating factor of the consequences of famine, as measured by diet quality. While interventions cannot retroactively add years of education to the target group, they can provide targeted nutritional education to those affected by famine. Those trainings are well established, and they include not only trained professionals discussing the importance of nutrition but they also include practical trainings in food preparation, including using nutritious local ingredients. Nutrition education is also a critical element in cash transfer interventions, to avoid use of resources for the purchase of low quality, energy dense foods.

In the population exposed to food shortage, those who were between 0–5 years of age when the shortage occurred are most severely affected, according to my findings. Therefore, in designing policy interventions, this group should have priority.

Results of my analysis suggest that those who survived famine and those who were born after a famine but in the affected areas have lower DDS scores, compared to those who were not exposed. The analysis also showed that the famine effects diminish over time. Additionally, a lower DDS of lactating women is found to hurt milk composition, suggesting that dietary consequences are being passed to next generation. Taken together, while famine survivors should have priority when it comes to post-famine support, their children and the population from the affected areas born after the famine should also be included in health monitoring activities.

7 Conclusion

7.1 Summary of the Dissertation

In this dissertation I analyze long-terms effects of famine on nutrient intake. To achieve that, in my empirical analysis I use several measures of famine and several measures of nutrient intake. Irrespective of measures used, I obtain similar results – famine has negative long-term effects on nutrient intake. To the best of my knowledge, this is the first attempt to establish the long-term relationship between famine and nutrient intake.

My findings are relevant for two main reasons:

Firstly, my findings reveal long-term negative consequences of famine on nutrient intake. Comparing to those who were never exposed to famine, diet of famine survivors is of poorer quality. Nutritionists, health practitioners and policy makers can benefit from these findings in the process of studying which factors influence individuals' eating patterns. In attempt to improve population's diet, nutritionists, health practitioners and policy makers can now use this evidence that famine survivors should receive more attention than those parts of the populations who were never exposed to famine. Therefore, a set of new or existing policy instruments should be targeted towards this population group.

Secondly, my findings contribute to the body of literature which explore determinants of eating patterns. While literature in the past focused on direct food consumption drivers, such as food price, access to food and food product characteristics, recent literature focused on underlying factors such as societal inequality or psychological determinants, such as mood, stress and guilt. My research contributes to body of research that deals with underlying determinants of food choice.

7.2 Caveats

My dissertation comes with several caveats.

In *Section One* and *Section Two*, the China Health and Nutrition Survey (CHNS) does not contain explicit information about whether a respondent was exposed to famine. Instead, I made several assumptions, which I elaborate on in *Methodology* section of *Chapter One*. Some studies argues that the Great Famine did not peak at the same time across and within provinces. Thus, this could have potentially impacted my analysis of the most critical year-age affected by famine. Additionally, CHNS is a repeated cross-sectional dataset, which does not lend itself to panel analysis that explicitly takes into account individual unobservable characteristics. A potential solution would be to use a (panel) dataset that tracks sae individuals over time, which would help alleviate a potential concern that unobservable individual characteristics might be driving dietary diversity scores.

In contrast to most existing studies, in my analysis I identify respondents who have not moved from their place of birth. This allows me to use province-level excess death rate EDR as a proxy for exposure to famine, assuming that respondents stayed in the same province as the province of their birth. An advantage of this approach is that it provides a less noisy measure of famine exposure relative to most existing literature. A potential drawback is that respondents who are more likely to stay within their province of birth are more likely to be male (since most women move in to live with their spouses upon marriage), and also located in rural areas (non-movers are less likely to be exposed to "flight to quality" of urban areas).

In Section *Two*, the estimated results in the analysis of differentiated effects of famine on nutrient intake by year cohort are obtained on relatively small sample sizes, (around 250 observations for cross-sectional regressions and 1064 for the pooled regression). Hence, the power of this analysis can be a potential concern.

When it comes to *Section Four*, in China Health and Retirement Longitudinal Study (CHARLS), respondents answered several crucial questions:

1. *When you were a child before age 17 was there ever a time when your family did not have enough food to eat?*
2. *Between 1958–1962 did you and your family (including your grandparents, parents, siblings, children and so on) experience starvation?*
3. *During those days, had any of your family (including your grandparents, parents, siblings, children and so on) starved to death?*

All three questions provide yes/no answers.

In Question 1, my analysis would benefit from deeper understanding of how respondent defines "enough food", and if it was only one family member who had to skip a meal or reduce regular food intake, or it was whole family who sacrificed. Perception of these two dimensions might have differed among respondents.

In Question 2, by applying this binary answer approach, I assume that, in the analysis of effects of starvation during the Great Famine on food expenditures, there is no difference whether respondents were starving themselves or it was their grandparents who starved, as the dataset does not provide such information.

In case of Question 3, there is a follow-up question that probes how many family members died due to starvation during the Great Famine. As in previous case, if a respond stated that one family member died due to starvation, I had to assume that there is no difference if that one person was respondent's child or grandparent.

7.3 Further Research

To address some of the caveats, I propose future data collection activities to investigate "famine experience" in detail. That data collection should happen soon after a famine episode has occurred, which will improve the accuracy of the information. Also, before answering questions, respondents should be informed about different degrees of food shortage. This would facilitate not only analysis of that famine but would also enable cross-country comparison as well as comparison of different food shortage episodes. To capture individual traits, future datasets should be longitudinal.

Further research could more deeply analyse the effects of exposure to different degrees of food shortages on food intake and qualitative characteristics of diets. This dissertation focuses on macronutrients and their relative share in diet. Future research could examine the effects of famine on the consumption of different types or sources of the same nutrient, and health implications associated with that. For example, research found that different sources of protein have different association with new-onset diabetes. There is U-shaped association between whole grain derived and poultry derived proteins and new-onset diabetes, while this association is J-shaped in case of red-meat derived protein, reverse J-shaped in case of fish-derived protein (Zhou et al., 2022). Therefore, future research might examine long-term impact of famine on intake of fish-derived proteins. The same applies to different sources of fat and carbohydrates.

In the same vein, several studies explore both DDS and food variety within a single food group and find that they affect health outcomes. Future research could investigate the relationship between exposure to famine and distribution of different food items within the same food group.

The long-term impact of exposure to famine on "food practices" is another potential area of research. By food practices, I refer to food purchasing patterns as well as cooking methods and intrahousehold allocation. Purchasing patterns are not limited to how often and where the food is being purchased. They also include information on who in a household decides what will be bought and who pays. The intrahousehold "food practices" dynamics could also include analyses of who decides what will be cooked, who will eat what, whether all household members eat at the same time, or whether there is order of who eats first, etc.

This dissertation showed that there is diminishing effect of famine over time. However, in my work I did not explore spatial dimension, which would analyse if there were diminishing effect as one moves away from "epicentre" of famine.

Additionally, future research might include analysis of long-term effects of famine on eating patterns the highest income groups. The findings could answer the question whether high income famine survivor expresses frugal eating behaviour or overcompensates food shortage from the past by eating excessive amounts of food when in position to do so.

Data Use Acknowledgements

This research uses data from China Health and Nutrition Survey (CHNS). We are grateful to research grant funding from the National Institute for Health (NIH), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) for R01 HD30880, National Institute on Aging (NIA) for R01 AG065357, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) for R01DK104371 and R01HL108427, the NIH Fogarty grant D43 TW009077 since 1989, and the China-Japan Friendship Hospital, Ministry of Health for support for CHNS 2009, Chinese National Human Genome Center at Shanghai since 2009, and Beijing Municipal Center for Disease Prevention and Control since 2011. We thank the National Institute for Nutrition and Health, China Center for Disease Control and Prevention, Beijing Municipal Center for Disease Control and Prevention, and the Chinese National Human Genome Center at Shanghai.

This analysis uses data or information from the Harmonized CHARLS dataset and Codebook, Version D as of June 2021 developed by the Gateway to Global Aging Data. The development of the Harmonized CHARLS was funded by the National Institute on Aging (R01 AG030153, RC2 AG036619, R03 AG043052). For more information, please refer to https://g2aging.org/.

8 References

Abrevaya, J. (1997). The equivalence of two estimators of the fixed-effects logit model. *Economics Letters*, *55*(1), 41–43. https://doi.org/10.1016/S0165-1765(97)00044-X

Almond, D., Edlund, L., Li, H., & Zhang, J. (2010). Long-Term Effects of Early-Life Development: Evidence from the 1959 to 1961 China Famine. In T. Ito & A. Rose (Eds.), *The Economic Consequences of Demographic Change in East Asia* (NBER, Vol. 19, pp. 321–345). University of Chicago Press.

Arage, G., Belachew, T., Abera, M., Abdulhay, F., Abdulahi, M., & Hassen Abate, K. (2020). Consequences of early life exposure to the 1983–1985 Ethiopian Great Famine on cognitive function in adults: a historical cohort study. *BMJ Open*, *10*(9), e038977. https://doi.org/10.1136/bmjopen-2020-038977

Babalola, D. A., & Isitor, S. U. (2014). Analysis of the Determinants of Food Expenditure Patterns among Urban Households in Nigeria: Evidence from Lagos State. *Journal of Agriculture and Veterinary Science*, *7*(5), 71–75.

Bauer, L., Broady, K., Edelberg, W., & O'Donnell, J. (2020). *Ten facts about COVID-19 and the U.S. economy*.

Benton, D. (2010). Neurodevelopment and neurodegeneration: are there critical stages for nutritional intervention? *Nutrition Reviews*, *68*, S6–S10. https://doi.org/10.1111/j.1753-4887.2010.00324.x

Ben-Zur, H., & Zeidner, M. (2009). Threat to Life and Risk-Taking Behaviors: A Review of Empirical Findings and Explanatory Models. *Personality and Social Psychology Review*, *13*(2), 109–128. https://doi.org/10.1177/1088868308330104

Bernile, G., Bhagwat, V., & Rau, P. R. (2017). What Doesn't Kill You Will Only Make You More Risk-Loving: Early-Life Disasters and CEO Behavior. *The Journal of Finance*, *72*(1), 167–206. https://doi.org/10.1111/jofi.12432

Bi, J., Liu, C., Li, S., He, Z., Chen, K., Luo, R., Wang, Z., Yu, Y., & Xu, H. (2019). Dietary Diversity among Preschoolers: A Cross-Sectional Study in Poor, Rural, and Ethnic Minority Areas of Central South China. *Nutrients*, *11*(3), 558. https://doi.org/10.3390/nu11030558

Birner, R., Blaschke, N., Bosch, C., Daum, T., Graf, S., Güttler, D., Heni, J., Kariuki, J., Katusiime, R., Seidel, A., Senon, Z. N., & Woode, G. (2021). 'We would rather die from Covid-19 than from hunger' - Exploring lockdown stringencies in five African countries. *Global Food Security*, *31*, 100571. https://doi.org/10.1016/j.gfs.2021.100571

Bucciol, A., & Zarri, L. (2015). The shadow of the past: Financial risk taking and negative life events. *Journal of Economic Psychology*, *48*, 1–16. https://doi.org/10.1016/j.joep.2015.02.006

Buchholz, K. (2022, April 12). *This is how wheat shortages are creating a food security risk*. World Economic Forum.

Carballo-Fazanes, A., Rico-Díaz, J., Barcala-Furelos, R., Rey, E., Rodríguez-Fernández, J. E., Varela-Casal, C., & Abelairas-Gómez, C. (2020). Physical Activity Habits and Determinants, Sedentary

Behaviour and Lifestyle in University Students. *International Journal of Environmental Research and Public Health*, *17*(9), 3272. https://doi.org/10.3390/ijerph17093272

CDC. (2020). *Micronutrient Facts*. https://www.cdc.gov/nutrition/micronutrient-malnutrition/micronutrients/index.html

Chadwick, R. (2013). Bioethics and lifestyle. *Bioethics*, *27*(7), ii–ii. https://doi.org/10.1111/bioe.12062

Chang, X., DeFries, R. S., Liu, L., & Davis, K. (2018). Understanding dietary and staple food transitions in China from multiple scales. *PLOS ONE*, *13*(4), e0195775. https://doi.org/10.1371/journal.pone.0195775

Chang, G. H., & Wen, G. J. (1997). Communal Dining and the Chinese Famine of 1958–1961. Economic Development and Cultural Change, 46(1), 1–34. https://doi.org/10.1086/452319

Chen, H., de la Rupelle, M., & Zilibotti, F. (2018). *The Great Famine and savings rate in China*.

Chen, Y., & Zhou, L.-A. (2007). The long-term health and economic consequences of the 1959–1961 famine in China. *Journal of Health Economics*, *26*(4), 659–681. https://doi.org/10.1016/j.jhealeco.2006.12.006

Cheng, Z., & Smyth, R. (2021). *Does Childhood Adversity Affect Household Portfolio Decisions? Evidence from the Chinese Great Famine*.

China CDC. (2004). *China Food Composition 2004* (Y. Yuexin, H. Mei, & P. Xingchang, Eds.). Peking University Medical Press.

Conklin, A. I., Monsivais, P., Khaw, K.-T., Wareham, N. J., & Forouhi, N. G. (2016). Dietary Diversity, Diet Cost, and Incidence of Type 2 Diabetes in the United Kingdom: A Prospective Cohort Study. *PLOS Medicine*, *13*(7), e1002085. https://doi.org/10.1371/journal.pmed.1002085

Cui, H., Smith, J. P., & Zhao, Y. (2020). Early-life deprivation and health outcomes in adulthood: Evidence from childhood hunger episodes of middle-aged and elderly Chinese. *Journal of Development Economics*, *143*, 102417. https://doi.org/10.1016/j.jdeveco.2019.102417

Dangura, D., & Gebremedhin, S. (2017). Dietary diversity and associated factors among children 6-23 months of age in Gorche district, Southern Ethiopia: Cross-sectional study. *BMC Pediatrics*, *17*(1), 6. https://doi.org/10.1186/s12887-016-0764-x

Devereux, S. (2000). *Famine in the Twentieth Century* (No. 105; IDS Working Papers).

Dieteren, C., & Bonfrer, I. (2021). Socioeconomic inequalities in lifestyle risk factors across low- and middle-income countries. *BMC Public Health*, *21*(1), 951. https://doi.org/10.1186/s12889-021-11014-1

Ding, Y., Min, S., Wang, X., & Yu, X. (2022). Memory of famine: The persistent impact of famine experience on food waste behavior. *China Economic Review*, *73*, 101795. https://doi.org/10.1016/j.chieco.2022.101795

Doan, D. (2014). *Does income growth improve diet diversity in China?*

Drescher, L., Thiele, S., Roosen, J., & Mensink, G. B. (2009). Consumer demand for healthy eating considering diversity - an economic approach for German individuals. *International Journal of Consumer Studies*, *33*(6), 684–696. https://doi.org/10.1111/j.1470-6431.2009.00812.x

Dunn, E. C., Soare, T. W., Zhu, Y., Simpkin, A. J., Suderman, M. J., Klengel, T., Smith, A. D. A. C., Ressler, K. J., & Relton, C. L. (2019). Sensitive Periods for the Effect of Childhood Adversity on DNA Methylation: Results From a Prospective, Longitudinal Study. *Biological Psychiatry*, *85*(10), 838–849. https://doi.org/10.1016/j.biopsych.2018.12.023

Edgerton-Tarpley, K. (2008). Tears from iron: cultural responses to famine in nineteenth-century China. University of California Press.

EUFIC. (2006, June 6). *The Factors That Influence Our Food Choices*.

Feng, X., & Johansson, A. C. (2018). Living through the Great Chinese Famine: Early-life experiences and managerial decisions. *Journal of Corporate Finance*, *48*, 638–657. https://doi.org/10.1016/j.jcorpfin.2017.11.012

Fung, W., & Ha, W. (2010). Intergenerational Effects of the 1959–61 China Famine. In *Risk, Shocks, and Human Development* (pp. 222–254). Palgrave Macmillan UK. https://doi.org/10.1057/9780230274129_10

Gabard-Durnam, L. J., & McLaughlin, K. A. (2019). Do Sensitive Periods Exist for Exposure to Adversity? *Biological Psychiatry*, *85*(10), 789–791. https://doi.org/10.1016/j.biopsych.2019.03.975

Ge, K., Zhai, F., & Wang, Q. (1997). Effect of nationality on dietary pattern and meal behavior in China. *The American Journal of Clinical Nutrition*, *65*(4), 1290S-1294S. https://doi.org/10.1093/ajcn/65.4.1290S

Global Diet Quality Project. (2021). *DQQ Tools & Data*. https://www.globaldietquality.org/home

Habib, H. S., Malik, A. M., Ali, A., & Khan, M. A. (2018). Socioeconomic determinants of rural household food expenditures in Rawalpindi. *Pakistan Journal of Agricultural Research*, *29*(1), 68–75.

Han, S., Zhang, H., Qin, L., & Zhai, C. (2013). Effects of Dietary Carbohydrate Replaced with Wild Rice (Zizania latifolia (Griseb) Turcz) on Insulin Resistance in Rats Fed with a High-Fat/Cholesterol Diet. *Nutrients*, *5*(2), 552–564. https://doi.org/10.3390/nu5020552

Haste, F. M., Brooke, O. G., Anderson, H. R., Bland, J. M., & Peacock, J. L. (1990). Social determinants of nutrient intake in smokers and non-smokers during pregnancy. *Journal of Epidemiology & Community Health*, *44*(3), 205–209. https://doi.org/10.1136/jech.44.3.205

Herman, D. R., Taylor Baer, M., Adams, E., Cunningham-Sabo, L., Duran, N., Johnson, D. B., & Yakes, E. (2014). Life Course Perspective: Evidence for the Role of Nutrition. *Maternal and Child Health Journal*, *18*(2), 450–461. https://doi.org/10.1007/s10995-013-1280-3

HLPE. (2020). *Food security and nutrition: building a global narrative towards 2030*. www.fao.org/cfs/cfs-hlpe

Hou, M., Qing, P., & Min, S. (2021). Multiple indicators of household dietary diversity in rural China: Effects of income and dietary knowledge. *Nutrition*, *91–92*, 111406. https://doi.org/10.1016/j.nut.2021.111406

Hu, J., Long, W., Tian, G. G., & Yao, T. (2017). CEOss Great Chinese Famine Experience and Accounting Conservatism: Evidence from China. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3107451

Hu, X. F., Liu, G. G., & Fan, M. (2017). Long-Term Effects of Famine on Chronic Diseases: Evidence from China's Great Leap Forward Famine. *Health Economics*, *26*(7), 922–936. https://doi.org/10.1002/hec.3371

IPCC. (2019). *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems* (P. R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, … J. Malley, Eds.).

Irwin, R. E., Pentieva, K., Cassidy, T., Lees-Murdock, D. J., McLaughlin, M., Prasad, G., McNulty, H., & Walsh, C. P. (2016). The interplay between DNA methylation, folate and neurocognitive development. *Epigenomics*, *8*(6), 863–879. https://doi.org/10.2217/epi-2016-0003

Jensen, M. (2007). Defining lifestyle. *Environmental Sciences*, *4*(2), 63–73. https://doi.org/10.1080/15693430701472747

Keister, L., Benton, R., & Moody, J. (2016). Lifestyles through Expenditures: A Case-Based Approach to Saving. *Sociological Science*, *3*, 650–684. https://doi.org/10.15195/v3.a28

Kim, Y.-I., & Lee, J. (2014). The long-run impact of a traumatic experience on risk aversion. *Journal of Economic Behavior & Organization*, *108*, 174–186. https://doi.org/10.1016/j.jebo.2014.09.009

Li, L. M. (2007). Fighting Famine in North China: State, Market, and Environmental Decline, 1690s-1990s. Stanford University Press.

Li, Q., & An, L. (2015). Intergenerational health consequences of the 1959–1961 Great Famine on children in rural China. *Economics & Human Biology*, *18*, 27–40. https://doi.org/10.1016/j.ehb.2015.03.003

Li, S., Chen, K., Liu, C., Bi, J., He, Z., Luo, R., Yu, Y., & Wang, Z. (2021). Association of dietary diversity and cognition in preschoolers in rural China. *Nutrition*, *91–92*, 111470. https://doi.org/10.1016/j.nut.2021.111470

Li, Y., He, Y., Qi, L., Jaddoe, V. W., Feskens, E. J. M., Yang, X., Ma, G., & Hu, F. B. (2010). Exposure to the Chinese Famine in Early Life and the Risk of Hyperglycemia and Type 2 Diabetes in Adulthood. *Diabetes*, *59*(10), 2400–2406. https://doi.org/10.2337/db10-0385

Li, Y., Jaddoe, V. W., Qi, L., He, Y., Lai, J., Wang, J., Zhang, J., Hu, Y., Ding, E. L., Yang, X., Hu, F. B., & Ma, G. (2011). Exposure to the Chinese famine in early life and the risk of hypertension in adulthood. *Journal of Hypertension*, *29*(6), 1085–1092. https://doi.org/10.1097/HJH.0b013e328345d969

Li, Y., Li, H., Decety, J., & Lee, K. (2013). Experiencing a Natural Disaster Alters Children's Altruistic Giving. *Psychological Science*, *24*(9), 1686–1695. https://doi.org/10.1177/0956797613479975

Lin, J. Y., & Yang, D. T. (1998). On the causes of China's agricultural crisis and the great leap famine. *China Economic Review*, *9*(2), 125–140. https://doi.org/10.1016/S1043-951X(99)80010-8

Liu, T. W. (2022). *Shanghai's Food Shortages Spur Voluntarism and Cynicism*.

Luo, Z., Mu, R., & Zhang, X. (2006). Famine and Overweight in China. *Review of Agricultural Economics*, *28*(3), 296–304. https://doi.org/10.1111/j.1467-9353.2006.00290.x

Lyu, S., Su, J., Xiang, Q., & Wu, M. (2014). Association of dietary pattern and physical activity level with triglyceride to high-density lipoprotein cholesterol ratio among adults in Jiangsu, China: a cross-sectional study with sex-specific differences. *Nutrition Research*, *34*(8), 674–681. https://doi.org/10.1016/j.nutres.2014.07.007

Malmendier, U., Tate, G., & Yan, J. (2011). Overconfidence and Early-Life Experiences: The Effect of Managerial Traits on Corporate Financial Policies. *The Journal of Finance*, *66*(5), 1687–1733. https://doi.org/10.1111/j.1540-6261.2011.01685.x

McPhee, J. S., French, D. P., Jackson, D., Nazroo, J., Pendleton, N., & Degens, H. (2016). Physical activity in older age: perspectives for healthy ageing and frailty. *Biogerontology*, *17*(3), 567–580. https://doi.org/10.1007/s10522-016-9641-0

Meng, X., Qian, N., & Yared, P. (2015). The Institutional Causes of China's Great Famine, 1959–1961. *The Review of Economic Studies*, *82*(4), 1568–1611. https://doi.org/10.1093/restud/rdv016

Moon, W., Florkowski, W. J., Beuchat, L. R., Resurreccion, A. v., Paraskova, P., Jordanov, J., & Chinnan, M. S. (2002). Demand for food variety in an emerging market economy. *Applied Economics*, *34*(5), 573–581. https://doi.org/10.1080/00036840110037863

Nguyen, P. H., Huybregts, L., Sanghvi, T. G., Tran, L. M., Frongillo, E. A., Menon, P., & Ruel, M. T. (2018). Dietary Diversity Predicts the Adequacy of Micronutrient Intake in Pregnant Adolescent Girls and Women in Bangladesh, but Use of the 5-Group Cutoff Poorly Identifies Individuals with Inadequate Intake. *The Journal of Nutrition*, *148*(5), 790–797. https://doi.org/10.1093/jn/nxy045

Ning, F., Ren, J., Song, X., Zhang, D., Liu, L., Zhang, L., Sun, J., Zhang, D., Pang, Z., Qiao, Q., & Diabetes Prevention Program, on behalf of Q. (2019). Famine Exposure in Early Life and Risk of Metabolic Syndrome in Adulthood: Comparisons of Different Metabolic Syndrome Definitions. *Journal of Diabetes Research*, *2019*, 1–9. https://doi.org/10.1155/2019/7954856

Nisbett, N., Harris, J., Backholer, K., Baker, P., Jernigan, V. B. B., & Friel, S. (2022). Holding no-one back: The Nutrition Equity Framework in theory and practice. *Global Food Security*, *32*, 100605. https://doi.org/10.1016/j.gfs.2021.100605

Oberg, C., Hodges, H., & Masten, A. S. (2021). Risk and Resilience of Somali Children in the Context of Climate Change, Famine, and Conflict. *Journal of Applied Research on Children: Informing Policy for Children at Risk*, *12*(1).

Ogle, C. M., Rubin, D. C., & Siegler, I. C. (2013). The impact of the developmental timing of trauma exposure on PTSD symptoms and psychosocial functioning among older adults. *Developmental Psychology*, *49*(11), 2191–2200. https://doi.org/10.1037/a0031985

Parpia, B. J. (1995). *Socioeconomic determinants of food and nutrient intakes in rural China*. Cornell University.

Pingali, P. (2007). Westernization of Asian diets and the transformation of food systems: Implications for research and policy. *Food Policy*, *32*(3), 281–298. https://doi.org/10.1016/j.foodpol.2006.08.001

Popkin, B. M. (1993). *Nutritional Patterns and Transitions* (Vol. 19, Issue 1). https://about.jstor.org/terms

Qin, L.-L., Luo, B.-A., Gao, F., Feng, X.-L., & Liu, J.-H. (2020). Effect of Exposure to Famine during Early Life on Risk of Metabolic Syndrome in Adulthood: A Meta-Analysis. *Journal of Diabetes Research*, *2020*, 1–9. https://doi.org/10.1155/2020/3251275

Qorbani, M., Mahdavi-Gorabi, A., Khatibi, N., Ejtahed, H.-S., Khazdouz, M., Djalalinia, S., Sahebkar, A., Esmaeili-Abdar, M., & Hasani, M. (2021). Dietary diversity score and cardio-metabolic risk factors: an updated systematic review and meta-analysis. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity*. https://doi.org/10.1007/s40519-020-01090-4

Rebello, S. A., Koh, H., Chen, C., Naidoo, N., Odegaard, A. O., Koh, W.-P., Butler, L. M., Yuan, J.-M., & van Dam, R. M. (2014). Amount, type, and sources of carbohydrates in relation to ischemic heart disease mortality in a Chinese population: a prospective cohort study. *The American Journal of Clinical Nutrition*, *100*(1), 53–64. https://doi.org/10.3945/ajcn.113.076273

Rong, H., Lai, X., Mahmoudi, E., & Fang, H. (2019). Early-Life Exposure to the Chinese Famine and Risk of Cognitive Decline. *Journal of Clinical Medicine*, *8*(4), 484. https://doi.org/10.3390/jcm8040484

Rubhara, T. T., Oduniyi, O. S., Mudhara, M., & Akwasi, A. M. (2020). Analysis of household food expenditure patterns. A case of Shamva district Zimbabwe. *Future of Food: Journal on Food, Agriculture and Society*, *8*(1).

Ruel, M. T. (2003). Is dietary diversity an indicator of food security or dietary quality? A review of measurement issues and research needs. *Food and Nutrition Bulletin*, *24*(2), 231–232. https://doi.org/10.1177/156482650302400210

Ruiz, M. (2021). Going beyond lifestyle factors. *ELife*, *10*. https://doi.org/10.7554/eLife.70548

Salehi-Abargouei, A., Akbari, F., Bellissimo, N., & Azadbakht, L. (2016). Dietary diversity score and obesity: a systematic review and meta-analysis of observational studies. *European Journal of Clinical Nutrition*, *70*(1), 1–9. https://doi.org/10.1038/ejcn.2015.118

Sekhampu, T. J. (2012). Socio-Economic Determinants of Household Food Expenditure in a Low Income Township in South Africa. *Mediterranean Journal of Social Sciences*, *3*(3).

Sen, A. (1981). Poverty and Famines: An Essay on Entitlement and Deprivation. Oxford University Press.

Sen, A. (2000). Development as Freedom. Anchor; Reprint edition (August 15, 2000).

Shi, Z., Zhang, C., Zhou, M., Zhen, S., & Taylor, A. W. (2013). Exposure to the Chinese famine in early life and the risk of anaemia in adulthood. *BMC Public Health*, *13*(1), 904. https://doi.org/10.1186/1471-2458-13-904

Slavin, P. (2016). Climate and famines: a historical reassessment. *WIREs Climate Change*, *7*(3), 433–447. https://doi.org/10.1002/wcc.395

Song, S., Wang, W., & Hu, P. (2009). Famine, death, and madness: Schizophrenia in early adulthood after prenatal exposure to the Chinese Great Leap Forward Famine. *Social Science & Medicine*, *68*(7), 1315–1321. https://doi.org/10.1016/j.socscimed.2009.01.027

Sotsha, K., Rambau, K., Khoza, T., Mmbengwa, V., & Ngqangweni, S. (2019). Socio-Economic Determinants of Rural Household Food Expenditure: A Quantile Regression Analysis. *OIDA International Journal of Sustainable Development*, *12*(2), 19–26.

Sparen, P., Vågerö, D., Shestov, D. B., Plavinskaja, S., Parfenova, N., Hoptiar, V., Paturot, D., & Galanti, M. R. (2004). Long term mortality after severe starvation during the siege of Leningrad: prospective cohort study. *BMJ*, *328*(7430), 11–0. https://doi.org/10.1136/bmj.37942.603970.9A

Tan, C. M., Zhibo, T., & Zhang, X. (2014). Sins of the Father: The Intergenerational Legacy of the 1959-61 Great Chinese Famine on Children's Cognitive Development. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2409452

Tao, C., Zhao, Q., Glauben, T., & Ren, Y. (2020). Does Dietary Diversity Reduce the Risk of Obesity? Empirical Evidence from Rural School Children in China. *International Journal of Environmental Research and Public Health*, *17*(21), 8122. https://doi.org/10.3390/ijerph17218122

Thorne-Lyman, A. L., Valpiani, N., Sun, K., Semba, R. D., Klotz, C. L., Kraemer, K., Akhter, N., de Pee, S., Moench-Pfanner, R., Sari, M., & Bloem, M. W. (2010). Household Dietary Diversity and Food Expenditures Are Closely Linked in Rural Bangladesh, Increasing the Risk of Malnutrition Due to the Financial Crisis. *The Journal of Nutrition*, *140*(1), 182S-188S. https://doi.org/10.3945/jn.109.110809

Tian, X., Wu, M., Zang, J., Zhu, Y., & Wang, H. (2017). Dietary diversity and adiposity in Chinese men and women: an analysis of four waves of cross-sectional survey data. *European Journal of Clinical Nutrition*, *71*(4), 506–511. https://doi.org/10.1038/ejcn.2016.212

Tian, X., Xu, X., Zhang, K., & Wang, H. (2017). Gender difference of metabolic syndrome and its association with dietary diversity at different ages. *Oncotarget*, *8*(43), 73568–73578. https://doi.org/10.18632/oncotarget.20625

Turner, C., Aggarwal, A., Walls, H., Herforth, A., Drewnowski, A., Coates, J., Kalamatianou, S., & Kadiyala, S. (2018). Concepts and critical perspectives for food environment research: A global framework with implications for action in low- and middle-income countries. *Global Food Security*, *18*, 93–101. https://doi.org/10.1016/j.gfs.2018.08.003

Turvey, C. G. (2019). Calamities and Conflict Affecting Rural China 1929–1933. In H. Hu, F. Zhong, & C. G. Turvey (Eds.), Chinese Agriculture in the 1930s. Palgrave Macmillan.

UN FAO. (2020a). *Food and Agriculture Statistics*. https://www.fao.org/food-agriculture-statistics/en/

UN FAO. (2020b). *Hunger and food insecurity*. https://www.fao.org/hunger/en/#:~:text=Hunger%20is%20an%20uncomfortable%20or,insufficient%20consumption%20of%20dietary%20energy.&text=For%20decades%2C%20FAO%20has%20used,be%20referred%20to%20as%20undernourishment.

UN OCHA. (2022, February 17). *Angola - Food insecurity and malnutrition*.

UN Security Council. (2021). *Already Up 20 Per Cent, Acute Hunger Driven by Conflict, Instability Risks Increasing Further Due to Climate Change, COVID-19, Secretary-General Warns Security Council*.

UN WHO. (2015, November 19). *Stunting in a nutshell*. https://www.who.int/news/item/19-11-2015-stunting-in-a-nutshell#:~:text=Stunting%20is%20the%20impaired%20growth,WHO%20Child%20Growth%20Standards%20median.

UN WHO. (2020). *Malnutrition*. https://www.who.int/health-topics/malnutrition#tab=tab_1

UNHCR. (2020). *Famine Explained: Definition, causes and facts*. https://www.unrefugees.org/news/famine-explained-definition-causes-and-facts/#What%20is%20famine

Updegraff, J. A., & Taylor, S. E. (2000). From Vulnerability to Growth: Positive and Negative Effects of Stressful Life Events. In *Loss and Trauma: General and Close Relationship Perspectives* (1st Edition, pp. 3–28). Brunner-Routledge.

van den Berg, G. J., Pinger, P. R., & Schoch, J. (2016). Instrumental Variable Estimation of the Causal Effect of Hunger Early in Life on Health Later in Life. *The Economic Journal*, *126*(591), 465–506. https://doi.org/10.1111/ecoj.12250

Variyam, J., Blaylock, J. R., & Smallwood, D. (1998). *USDA's Healthy Eating Index and Nutrition Information*.

Veal, A. J. (1993). The concept of lifestyle: a review. *Leisure Studies*, *12*(4), 233–252. https://doi.org/10.1080/02614369300390231

Vyncke, P. (2002). Lifestyle Segmentation: From Attitudes, Interests and Opinions, to Values, Aesthetic Styles, Life Visions and Media Preferences. *European Journal of Communication*, *17*(4), 445–463. https://doi.org/10.1177/02673231020170040301

Wang, Z., Chen, Y., Tang, S., Chen, S., Gong, S., Jiang, X., Wang, L., & Zhang, Y. (2021). Dietary Diversity and Nutrient Intake of Han and Dongxiang Smallholder Farmers in Poverty Areas of Northwest China. *Nutrients*, *13*(11), 3908. https://doi.org/10.3390/nu13113908

WHO/UNICEF/WFP. (2014). *Global nutrition targets 2025: wasting policy brief (WHO/NMH/NHD/14.8)*.

World Food Programme. (n.d.). *The WFP food basket*.

World Food Programme. (2022a). *Yemen emergency*. https://www.wfp.org/emergencies/yemen-emergency

World Food Programme. (2022b, March 31). *Afghanistan*. https://www.wfp.org/countries/afghanistan

World Health Organization. (2014). *Global Nutrition Targets 2025: Stunting Policy Brief*.

Xu, H., Li, L., Zhang, Z., & Liu, J. (2016). Is natural experiment a cure? Re-examining the long-term health effects of China's 1959–1961 famine. *Social Science & Medicine*, *148*, 110–122. https://doi.org/10.1016/j.socscimed.2015.11.028

Xu, Y., Zhu, S., Zhang, T., Wang, D., Hu, J., Gao, J., & Zhou, Z. (2020). Explaining Income-Related Inequalities in Dietary Knowledge: Evidence from the China Health and Nutrition Survey. *International Journal of Environmental Research and Public Health*, *17*(2), 532. https://doi.org/10.3390/ijerph17020532

Yao, S., Wang, Z., Sun, M., Liao, J., & Cheng, F. (2020). Top executives' early-life experience and financial disclosure quality: impact from the Great Chinese Famine. *Accounting & Finance*, *60*(5), 4757–4793. https://doi.org/10.1111/acfi.12659

Yin, Z., Fei, Z., Qiu, C., Brasher, M. S., Kraus, V. B., Zhao, W., Shi, X., & Zeng, Y. (2017). Dietary diversity and cognitive function among elderly people: A population-based study. *The Journal of Nutrition, Health & Aging*, *21*(10), 1089–1094. https://doi.org/10.1007/s12603-017-0912-5

You, J., Imai, K. S., & Gaiha, R. (2016). Declining Nutrient Intake in a Growing China: Does Household Heterogeneity Matter? *World Development*, *77*, 171–191. https://doi.org/10.1016/j.worlddev.2015.08.016

Zani, M., Saediman, H., Abdullah, S., Daud. L, & Yunus, L. (2019). Determinants of Household Food Expenditure in a Cassava Growing Village in Southeast Sulawesi. *Academic Journal of Interdisciplinary Studies*, *8*(3).

Zhai, F. Y., Du, S. F., Wang, Z. H., Zhang, J. G., Du, W. W., & Popkin, B. M. (2014). Dynamics of the Chinese diet and the role of urbanicity, 1991-2011. *Obesity Reviews*, *15*, 16–26. https://doi.org/10.1111/obr.12124

Zhang, J., Liang, D., & Zhao, A. (2020). Dietary Diversity and the Risk of Fracture in Adults: A Prospective Study. *Nutrients*, *12*(12), 3655. https://doi.org/10.3390/nu12123655

Zhang, J., Wang, Z., Wang, H., Du, W., Su, C., Zhang, J., Jiang, H., Jia, X., Huang, F., Zhai, F., & Zhang, B. (2016). Association between dietary patterns and blood lipid profiles among Chinese women. *Public Health Nutrition*, *19*(18), 3361–3368. https://doi.org/10.1017/S136898001600197X

Zhang, J., & Zhao, A. (2021). Dietary Diversity and Healthy Aging: A Prospective Study. *Nutrients*, *13*(6), 1787. https://doi.org/10.3390/nu13061787

Zhang, J., Zhao, A., Wu, W., Yang, C., Ren, Z., Wang, M., Wang, P., & Zhang, Y. (2020). Dietary Diversity Is Associated With Memory Status in Chinese Adults: A Prospective Study. *Frontiers in Aging Neuroscience*, *12*. https://doi.org/10.3389/fnagi.2020.580760

Zhang, L. (2017). CEOs' early-life experiences and corporate policy: Evidence from China's great famine. *Pacific-Basin Finance Journal*, *46*, 57–77. https://doi.org/10.1016/j.pacfin.2017.08.004

Zhang, Q., Chen, X., Liu, Z., Varma, D. S., Wan, R., & Zhao, S. (2017). Diet diversity and nutritional status among adults in southwest China. *PLOS ONE*, *12*(2), e0172406. https://doi.org/10.1371/journal.pone.0172406

Zhao, A., Li, Z., Ke, Y., Huo, S., Ma, Y., Zhang, Y., Zhang, J., & Ren, Z. (2020). Dietary Diversity among Chinese Residents during the COVID-19 Outbreak and Its Associated Factors. *Nutrients*, *12*(6), 1699. https://doi.org/10.3390/nu12061699

Zhao, W., Yu, K., Tan, S., Zheng, Y., Zhao, A., Wang, P., & Zhang, Y. (2017). Dietary diversity scores: an indicator of micronutrient inadequacy instead of obesity for Chinese children. *BMC Public Health*, *17*(1), 440. https://doi.org/10.1186/s12889-017-4381-x

Zhao, Y., Hu, Y., Smith, J. P., Strauss, J., & Yang, G. (2014). Cohort Profile: The China Health and Retirement Longitudinal Study (CHARLS). *International Journal of Epidemiology*, *43*(1), 61–68. https://doi.org/10.1093/ije/dys203

Zhen, S., Ma, Y., Zhao, Z., Yang, X., & Wen, D. (2018). Dietary pattern is associated with obesity in Chinese children and adolescents: data from China Health and Nutrition Survey (CHNS). *Nutrition Journal*, *17*(1), 68. https://doi.org/10.1186/s12937-018-0372-8

Zhong, T., Si, Z., Crush, J., Xu, Z., Huang, X., Scott, S., Tang, S., & Zhang, X. (2018). The Impact of Proximity to Wet Markets and Supermarkets on Household Dietary Diversity in Nanjing City, China. *Sustainability*, *10*(5), 1465. https://doi.org/10.3390/su10051465

Zhou, C., Liu, C., Zhang, Z., Liu, M., Zhang, Y., Li, H., He, P., Li, Q., & Qin, X. (2022). Variety and quantity of dietary protein intake from different sources and risk of new-onset diabetes: a Nationwide Cohort Study in China. *BMC Medicine*, *20*(1), 6. https://doi.org/10.1186/s12916-021-02199-8

Annex

Table A. 1 Estimated mortality in major 20th century famines

Years	Location (epicentre)	Excess mortality	Causal triggers
1903-06	Nigeria (Hausa land)	5,000	Drought
1906-07	Tanzania (south)	37,500	Conflict
1913-14	West Africa (Sahel)	125,000	Drought
1917-19	Tanzania (central)	30,000	Conflict & Drought
1920-21	China (Gansu, Shaanxi)	500,000	Drought
1921-22	Soviet Union	**9,000,000**	Conflict & Drought
1927	China (northwest)	**3,000,000-6,000,000**	Natural disasters
1929	China (Hunan)	**2,000,000**	Conflict & Drought
1932-34	Soviet Union (Ukraine)	**7,000,000-8,000,000**	Government policy
1943	China (Henan)	**5,000,000**	Conflict
1943	India (Bengal)	**2,100,000-3,000,000**	Conflict
1943-44	Rwanda	300,000	Conflict & Drought
1944	Netherlands	10,000	Conflict
1946-47	Soviet Union	**2,000,000**	Drought & Government policy
1957-58	Ethiopia (Tigray)	100,000-397,000	Drought & Locusts
1958-62	China	**30,000,000-33,000,000**	Government policy
1966	Ethiopia (Wollo)	45,000-60,000	Drought
1968-70	Nigeria (Biafra)	1,000,000	Conflict
1969-74	West Africa (Sahel)	101,000	Drought
1972-73	India (Maharashtra)	130,000	Drought
1972-75	Ethiopia (Wollo & Tigray)	200,000-500,000	Drought
1974-75	Somalia	20,000	Drought & Government policy
1974	Bangladesh	**1,500,000**	Flood & market failure
1979	Cambodia	**1,500,000-2,000,000**	Conflict
1980-81	Uganda (Karamoja)	30,000	Conflict & Drought
1982-85	Mozambique	100,000	Conflict & Drought
1983-85	Ethiopia	590,000-1,000,000	Conflict & Drought
1984-85	Sudan (Darfur, Kordofan)	250,000	Drought
1988	Sudan (south)	250,000	Conflict
1991-93	Somalia	300,000-500,000	Conflict & Drought
1995-99	North Korea	**2,800,000-3,500,000**	Flood & Government policy
1998	Sudan (Bahr el Ghazal)	70,000	Conflict & Drought

Source: (Devereux, 2000)

Definitions 1: Definitions of different forms of food shortages, and related nutritional outcomes

Hunger – an uncomfortable or painful physical sensation caused by insufficient consumption of dietary energy (UN FAO, 2020b).

Prevalence of Undernourishment *(PoU)* estimates the adequacy of a population's dietary energy, and it is based on food availability, food consumption and energy needs (UN FAO, 2020b). In other words, it measures the amount of calories consumed against minimum physiological needs. Historically FAO used PoU as a measure of hunger.

Food Security exists when all people, at all times, have physical, social and economic access to sufficient, safe and nutritious food which meets their dietary needs and food preferences for an active and healthy life (UN FAO, 2020a). When food security does not exist, there can be moderate or severe food insecurity, which is based on Food Insecurity Experience Scale. In addition to traditional 4 elements of food security – food availability, access, utilization and stability, recent efforts added two more components, and these are agency and sustainability (HLPE, 2020).

Stunting – the impaired growth and development that children experience from poor nutrition, repeated infection, and inadequate psychosocial stimulation. Children are defined as stunted if their *height-for-age* is more than two standard deviations below the WHO Child Growth Standards median (UN WHO, 2015). Stunting is often seen as an indicator of chronic hunger.

Wasting – low *weight-for-height*. It often indicates recent and severe weight loss, although it can also persist for a long time. It usually occurs when a person has not had food of adequate quality and quantity and/or they have had frequent or prolonged illnesses (UN WHO, 2020).

Underweight – low *weight-for-age*. A child who is underweight may be stunted, wasted or both (UN WHO, 2020).

Micronutrient deficiency – lack of vitamins and minerals which are vital to healthy development, disease prevention, and wellbeing. They are not produced in the body and must be derived from the diet (CDC, 2020). Some of these are: iron, vitamin A, vitamin D, Iodine, Folate (vitamin B9) and Zinc, where iron deficiency (anaemia) is the most common form of micronutrient malnutrition globally

Famine is declared when certain levels of mortality, malnutrition and hunger are reached. They are: at least 20 per cent of households in an area face extreme food shortages with a limited ability to cope; acute malnutrition rates exceed 30 per cent; and the death rate exceeds two persons per day per 10,000 persons (UNHCR, 2020).

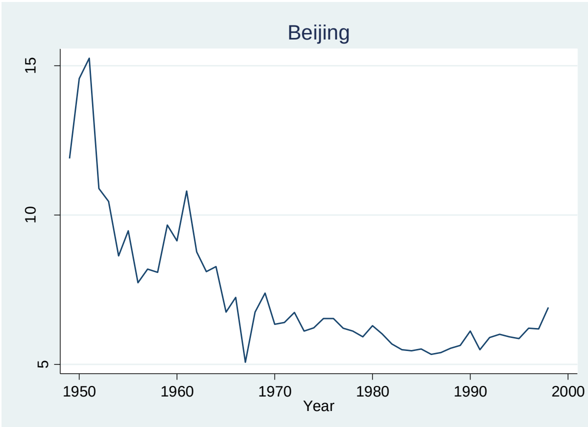
Malnutrition **-** refers to deficiencies or excesses in nutrient intake, imbalance of essential nutrients or impaired nutrient utilization (UN WHO, 2020). Coexistence of undernutrition and overweight at a certain location is called double burden of malnutrition, and when it also includes micronutrient deficiency, it is called triple burden of malnutrition.

Annex1

Table A.1. 1: Sample size in 10 waves of China Health and Nutrition Survey

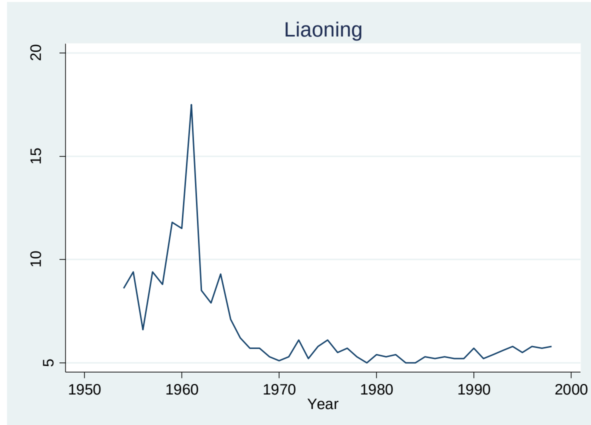
Survey Year	Number of Communities	Number of Households	Number of Individuals
1989	180	3,795	15,907
1991	189	3,619	14,797
1993	181	3,456	13,895
1997	191	3,875	14,441
2000	215	4,396	15,831
2004	216	4,387	12,308
2006	218	4,467	11,860
2009	217	4,517	12,178
2011	289	5,923	15,725
2015	360	7,319	20,914
Participated ever	388	11,130	42,829

Graph A.1 1: Death rate in Beijing



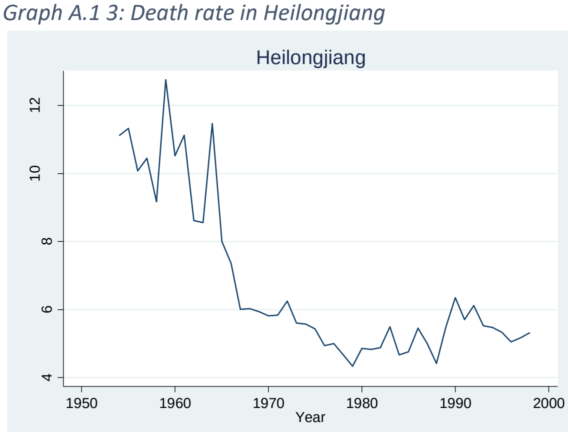
Source: Meng et al. (2009)

Graph A.1 2: Death rate in Liaoning



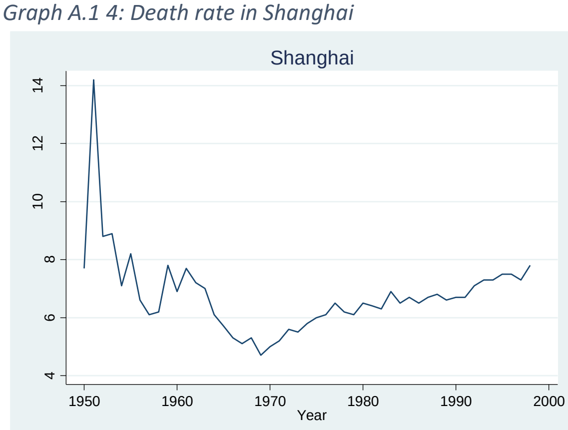
Source: Meng et al. (2009)

Graph A.1 3: Death rate in Heilongjiang



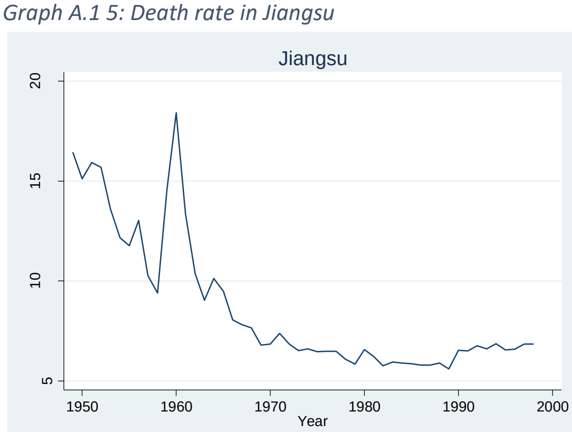
Source: Meng et al. (2009)

Graph A.1 4: Death rate in Shanghai



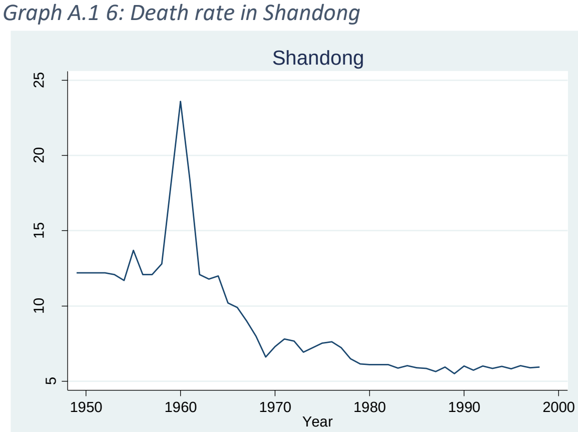
Source: Meng et al. (2009)

Graph A.1 5: Death rate in Jiangsu



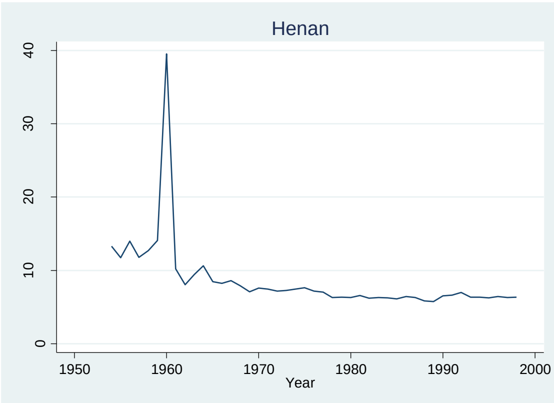
Source: Meng et al. (2009)

Graph A.1 6: Death rate in Shandong



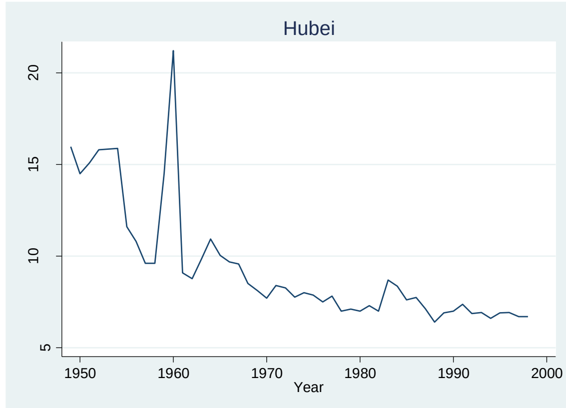
Source: Meng et al. (2009)

Graph A.1 7: Death rate in Henan



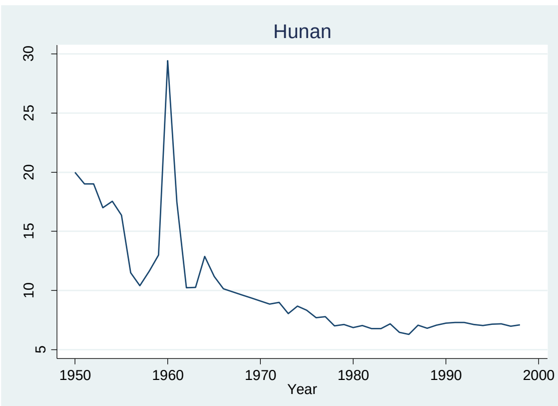
Source: Meng et al. (2009)

Graph A.1 8: Death rate in Hubei



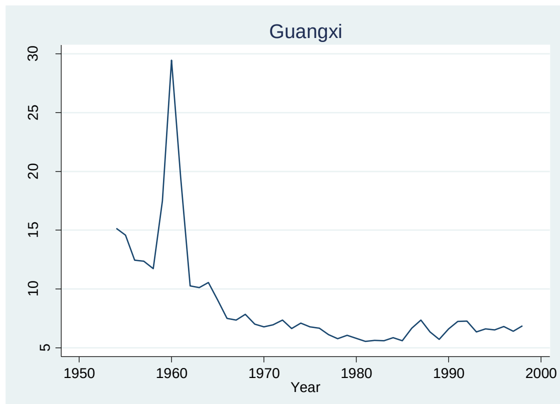
Source: Meng et al. (2009)

Graph A.1 9: Death rate in Hunan



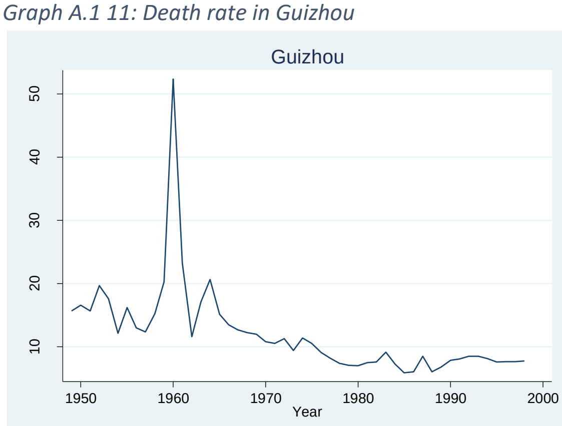
Source: Meng et al. (2009)

Graph A.1 10: Death rate in Guangxi



Source: Meng et al. (2009)

Graph A.1 11: Death rate in Guizhou



Source: Meng et al. (2009)

Table A.1. 2: CFCT 1991

Food Composition Table 1991 ('97, '00)	
A01	Cereals and Cereal Products
A02	Dried Legumes and Legume Products
A03	Fresh and Sprouted Legumes
A04	Roots, Tubers and Stems
A05	Fresh Leafy Vegetables
A06	Melons, Squashes and Gourds
A07	Fruit Bearing Vegetables
A08	Pickled, Salted and Preserved Vegetables
A09	Fungi and Algae
A10	Fruits
A11	Nuts and Seeds
A12	Meat
A13	Poultry
A14	Milk and Milk Products
A15	Infant Foods
A16	Eggs, Egg Products
A17	Fishes
A18	Mollusks and Invertebrates
A19	Crustaceans
A20	Fats and Oils
A21	Ethnic Food and Cakes
A22	Beverages
A23	Liquor and Alcoholic Beverages
A24	Confectionary
A25	Starch
A26	Condiments
A27	Spices
A28	Reptiles

Table A.1. 3 CFCT 2002/2004

Food Composition Table 2002/04 ('04, '06, '09, '11)	
01	Cereals and Cereal products
02	Tubers, Starches and Products
03	Dried Legumes and Legume Products
04	Vegetable and Vegetable Products
05	Fungi and Algae
06	Fruit and Fruit Products
07	Nuts and Seeds
08	Meat and Meat Products
09	Poultry and Poultry Products
10	Milk and Milk Products
11	Eggs and Egg Products
12	Fish, Shellfish, Mollusks
13	Infant Foods
14	Ethnic Food and Cakes
15	Fast Foods
16	Beverages
17	Liquor and Alcoholic Beverages
18	Sugars and Preserves
19	Fats and Oils
20	Condiments

Table A.1. 4: Provincial Excess Death Rate
EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years.

Province	EDR [%]	EDR [dif]
Beijing	1.23	1.87
Liaoning	1.65	5.33
Heilongjiang	1.16	1.57
Shanghai	1.19	1.17
Jiangsu	1.42	4.54
Shandong	1.63	7.73
Henan	1.66	8.46
Hubei	1.49	4.92
Hunan	1.78	8.77
Guangxi	1.82	9.97
Guizhou	2.36	18.42

Source: Meng et al. (2015)

Table A.1. 5: Effects of famine (EDR [%]) on Dietary Diversity Score; Cross sectional regressions

Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); Dependent variable Dietary Diversity Score is calculated using 20 food groups.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets.*

	Dietary Diversity Score			
	2004	2006	2009	2011
log (HH total income)	0.076***	0.065***	0.043***	0.063***
	[9.489]	[7.299]	[5.850]	[8.599]
Education	0.001	0.003***	0.002*	0.001
	[1.133]	[2.660]	[1.670]	[0.399]
HH size	-0.009***	-0.009***	-0.005	-0.002
	[-3.728]	[-2.926]	[-1.594]	[-0.459]
EDR [%]	-0.064***	-0.085***	-0.093***	-0.087***
	[-5.134]	[-6.355]	[-5.690]	[-4.982]
Gender	-0.002	-0.014	0.020	0.003
	[-0.203]	[-0.837]	[1.219]	[0.182]
Age	0.010	0.096**	0.009	-0.112**
	[0.277]	[2.087]	[0.166]	[-1.985]
Han nationality	-0.014	0.002	-0.023**	-0.038***
	[-1.541]	[0.235]	[-2.148]	[-3.256]
Old home food habits	0.006	0.025	-0.017	0.034
	[0.343]	[1.355]	[-0.828]	[1.257]
Marital Status	0.060***	0.037**	0.011	0.029
	[4.759]	[2.052]	[0.480]	[1.284]
Urban	0.034***	0.018	0.049**	0.033
	[3.093]	[1.107]	[2.449]	[1.640]
Observations	1,001	832	764	686
R-squared	0.163	0.176	0.127	0.187

Table A.1. 6: Effects of famine (EDR [diff]) on Dietary Diversity Score; Cross sectional regressions

Variables are defined as described in Table 1.2. EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using 20 food groups
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets.*

	Dietary Diversity Score			
	2004	2006	2009	2011
log (HH total income)	0.076***	0.065***	0.042***	0.063***
	[9.560]	[7.332]	[5.819]	[8.467]
Education	0.001	0.003**	0.002	0.001
	[1.012]	[2.497]	[1.460]	[0.461]
HH size	-0.008***	-0.008***	-0.003	-0.001
	[-3.530]	[-2.654]	[-1.221]	[-0.406]
EDR [diff]	-0.005***	-0.006***	-0.007***	-0.005***
	[-5.663]	[-7.409]	[-7.079]	[-4.795]
Gender	-0.002	-0.014	0.019	0.003
	[-0.178]	[-0.875]	[1.129]	[0.205]
Age	0.012	0.098**	0.011	-0.105*
	[0.334]	[2.151]	[0.203]	[-1.855]
Han nationality	-0.015	0.001	-0.026**	-0.033***
	[-1.636]	[0.136]	[-2.557]	[-2.840]
Old home food habits	0.008	0.028	-0.013	0.036
	[0.457]	[1.586]	[-0.655]	[1.355]
Marital Status	0.059***	0.034*	0.009	0.029
	[4.696]	[1.867]	[0.394]	[1.298]
Urban	0.034***	0.019	0.051**	0.034*
	[3.177]	[1.193]	[2.574]	[1.678]
Observations	1,001	832	764	686
R-squared	0.168	0.186	0.142	0.186

Table A.1. 7: Effects of famine (EDR [%]) on Dietary Diversity Score; Age convexity, Cross sectional regressions

Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); Dependent variable Dietary Diversity Score is calculated using 20 food groups.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets.*

	Dietary Diversity Score			
	2004	2006	2009	2011
log (HH total income)	0.074***	0.065***	0.043***	0.063***
	[9.407]	[7.261]	[5.840]	[8.584]
Education	0.001	0.003***	0.002*	0.000
	[1.074]	[2.642]	[1.687]	[0.337]
HH size	-0.010***	-0.010***	-0.005*	-0.002
	[-4.001]	[-3.257]	[-1.721]	[-0.650]
EDR [%]	-0.062***	-0.081***	-0.091***	-0.085***
	[-5.001]	[-5.967]	[-5.575]	[-4.773]
Gender	-0.002	-0.013	0.020	0.004
	[-0.153]	[-0.773]	[1.232]	[0.254]
Age	0.469*	0.731*	0.365	0.372
	[1.708]	[1.782]	[0.750]	[0.622]
Age^2	-0.410*	-0.551	-0.298	-0.393
	[-1.720]	[-1.583]	[-0.736]	[-0.805]
Han nationality	-0.013	0.004	-0.021**	-0.037***
	[-1.324]	[0.462]	[-2.015]	[-3.044]
Old home food habits	0.006	0.024	-0.017	0.034
	[0.326]	[1.349]	[-0.818]	[1.278]
Marital Status	0.059***	0.037**	0.010	0.028
	[4.573]	[2.067]	[0.398]	[1.237]
Urban	0.034***	0.017	0.049**	0.033*
	[3.096]	[1.059]	[2.441]	[1.663]
Observations	1,001	832	764	686
R-squared	0.165	0.179	0.128	0.188

Table A.1. 8: Effects of famine (EDR [diff]) on Dietary Diversity Score; Age convexity, Cross sectional regressions

Variables are defined as described in Table 1.2. EDR [diff] presents the difference between 3 famine years and 3 pre-famine years.; Dependent variable Dietary Diversity Score is calculated using 20 food groups.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets.*

	Dietary Diversity Score			
	2004	2006	2009	2011
log (HH total income)	0.074***	0.065***	0.042***	0.063***
	[9.470]	[7.289]	[5.810]	[8.454]
Education	0.001	0.003**	0.002	0.001
	[0.948]	[2.479]	[1.478]	[0.384]
HH size	-0.009***	-0.009***	-0.004	-0.002
	[-3.827]	[-2.991]	[-1.355]	[-0.625]
EDR [diff]	-0.004***	-0.006***	-0.007***	-0.005***
	[-5.567]	[-7.081]	[-6.992]	[-4.621]
Gender	-0.001	-0.013	0.019	0.004
	[-0.130]	[-0.817]	[1.141]	[0.284]
Age	0.486*	0.723*	0.355	0.444
	[1.772]	[1.776]	[0.739]	[0.747]
Age^2	-0.424*	-0.542	-0.288	-0.446
	[-1.776]	[-1.569]	[-0.719]	[-0.918]
Han nationality	-0.013	0.003	-0.025**	-0.031***
	[-1.439]	[0.338]	[-2.441]	[-2.658]
Old home food habits	0.008	0.028	-0.013	0.037
	[0.438]	[1.582]	[-0.645]	[1.379]
Marital Status	0.058***	0.034*	0.008	0.028
	[4.509]	[1.881]	[0.316]	[1.244]
Urban	0.034***	0.018	0.051**	0.034*
	[3.180]	[1.146]	[2.567]	[1.704]
Observations	1,001	832	764	686
R-squared	0.170	0.188	0.143	0.187

Table A.1. 9: Effects of interaction of Famine (EDR [%]) and Age on Dietary Diversity Score; Cross sectional regressions

Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); Dependent variable Dietary Diversity Score is calculated using 20 food groups.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets.*

	Dietary Diversity Score			
	2004	2006	2009	2011
log (HH total income)	0.075***	0.065***	0.043***	0.062***
	[9.454]	[7.310]	[5.880]	[8.498]
Education	0.001	0.003***	0.002*	0.001
	[1.118]	[2.638]	[1.711]	[0.422]
HH size	-0.009***	-0.009***	-0.005	-0.001
	[-3.805]	[-2.955]	[-1.568]	[-0.413]
EDR [%]	-0.156***	-0.146**	0.036	-0.276***
	[-2.850]	[-2.327]	[0.406]	[-2.712]
Age	-0.281	-0.094	0.378	-0.641**
	[-1.596]	[-0.449]	[1.388]	[-2.197]
EDR [%] x Age	0.166*	0.108	-0.212	0.300*
	[1.736]	[0.994]	[-1.447]	[1.843]
Gender	-0.001	-0.013	0.019	0.004
	[-0.108]	[-0.767]	[1.169]	[0.238]
Han nationality	-0.014	0.003	-0.023**	-0.038***
	[-1.477]	[0.350]	[-2.176]	[-3.226]
Old home food habits	0.007	0.024	-0.017	0.033
	[0.356]	[1.319]	[-0.807]	[1.232]
Marital Status	0.063***	0.039**	0.009	0.029
	[4.987]	[2.127]	[0.379]	[1.296]
Urban	0.036***	0.019	0.047**	0.035*
	[3.250]	[1.146]	[2.317]	[1.742]
Observations	1,001	832	764	686
R-squared	0.165	0.177	0.129	0.191

Table A.1. 10: Effects of interaction of Famine (EDR [diff]) and Age on Dietary Diversity Score; Cross sectional regressions

Variables are defined as described in Table 1.2. *EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using 20 food groups.*
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets.*

	Dietary Diversity Score			
	2004	2006	2009	2011
log (HH total income)	0.075***	0.065***	0.042***	0.062***
	[9.557]	[7.352]	[5.825]	[8.403]
Education	0.001	0.003**	0.002	0.001
	[1.018]	[2.491]	[1.457]	[0.522]
HH size	-0.008***	-0.008***	-0.003	-0.001
	[-3.617]	[-2.676]	[-1.195]	[-0.383]
EDR [diff]	-0.011***	-0.010**	-0.003	-0.016**
	[-3.119]	[-2.480]	[-0.532]	[-2.520]
Age	-0.090	0.041	0.071	-0.260**
	[-1.330]	[0.469]	[0.677]	[-2.414]
EDR [diff] x Age	0.012*	0.006	-0.007	0.017*
	[1.890]	[0.929]	[-0.759]	[1.668]
Gender	-0.001	-0.013	0.018	0.004
	[-0.078]	[-0.811]	[1.103]	[0.239]
Han nationality	-0.014	0.002	-0.026***	-0.031***
	[-1.517]	[0.267]	[-2.587]	[-2.753]
Old home food habits	0.009	0.027	-0.013	0.036
	[0.472]	[1.551]	[-0.643]	[1.333]
Marital Status	0.062***	0.035*	0.008	0.030
	[4.943]	[1.932]	[0.337]	[1.345]
Urban	0.036***	0.020	0.050**	0.036*
	[3.357]	[1.230]	[2.486]	[1.777]
Observations	1,001	832	764	686
R-squared	0.170	0.186	0.143	0.189

Table A.1. 11: Effects of interaction of Famine (EDR [%]) and Convex Age term on Dietary Diversity Score; Cross sectional regressions

Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); Dependent variable Dietary Diversity Score is calculated using 20 food groups.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets.*

	Dietary Diversity Score			
	2004	2006	2009	2011
log (HH total income)	0.074***	0.065***	0.043***	0.062***
	[9.343]	[7.271]	[5.874]	[8.455]
Education	0.001	0.003***	0.002*	0.000
	[1.045]	[2.609]	[1.726]	[0.318]
HH size	-0.010***	-0.010***	-0.005	-0.003
	[-4.151]	[-3.323]	[-1.632]	[-0.717]
EDR [%]	-0.124***	-0.119***	-0.032	-0.199***
	[-4.105]	[-3.554]	[-0.658]	[-3.659]
Age	0.555**	0.790*	0.247	0.662
	[2.025]	[1.911]	[0.486]	[1.097]
Age^2	-0.827***	-0.806*	0.081	-1.133*
	[-2.699]	[-1.956]	[0.150]	[-1.920]
EDR [%] x Age^2	0.194**	0.115	-0.160	0.285**
	[2.288]	[1.222]	[-1.253]	[2.128]
Gender	-0.000	-0.011	0.019	0.005
	[-0.032]	[-0.682]	[1.183]	[0.344]
Han nationality	-0.012	0.006	-0.022**	-0.035***
	[-1.244]	[0.599]	[-2.054]	[-2.965]
Old home food habits	0.007	0.023	-0.017	0.033
	[0.350]	[1.305]	[-0.807]	[1.265]
Marital Status	0.062***	0.039**	0.008	0.028
	[4.825]	[2.164]	[0.333]	[1.224]
Urban	0.036***	0.018	0.047**	0.036*
	[3.315]	[1.108]	[2.313]	[1.821]
Observations	1,001	832	764	686
R-squared	0.168	0.180	0.130	0.193

Table A.1. 12: Effects of interaction of Famine (EDR [diff]) and convex Age term on Dietary Diversity Score; Cross sectional regressions

Variables are defined as described in Table 1.2. *EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using 20 food groups.*
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets.*

	Dietary Diversity Score			
	2004	2006	2009	2011
log (HH total income)	0.074***	0.065***	0.042***	0.062***
	[9.443]	[7.311]	[5.818]	[8.363]
Education	0.001	0.003**	0.002	0.001
	[0.945]	[2.465]	[1.477]	[0.410]
HH size	-0.010***	-0.009***	-0.004	-0.003
	[-3.992]	[-3.050]	[-1.298]	[-0.715]
EDR [diff]	-0.009***	-0.008***	-0.005*	-0.012***
	[-4.475]	[-3.965]	[-1.770]	[-3.527]
Age	0.577**	0.784*	0.296	0.739
	[2.113]	[1.907]	[0.591]	[1.228]
Age^2	-0.625**	-0.659*	-0.198	-0.842
	[-2.481]	[-1.814]	[-0.445]	[-1.635]
EDR [diff] x Age^2 (Head)	0.014**	0.007	-0.005	0.017**
	[2.467]	[1.188]	[-0.581]	[2.006]
Gender	0.000	-0.012	0.018	0.005
	[0.001]	[-0.731]	[1.118]	[0.351]
Han nationality	-0.012	0.005	-0.025**	-0.029**
	[-1.288]	[0.502]	[-2.468]	[-2.501]
Old home food habits	0.009	0.027	-0.013	0.036
	[0.465]	[1.537]	[-0.639]	[1.368]
Marital Status	0.061***	0.036**	0.007	0.028
	[4.787]	[1.970]	[0.284]	[1.270]
Urban	0.037***	0.019	0.050**	0.037*
	[3.429]	[1.192]	[2.483]	[1.863]
Observations	1,001	832	764	686
R-squared	0.174	0.189	0.143	0.192

Table A.1. 13: Effects of famine (EDR [%/diff]) on Dietary Diversity Score; Pooled regressions (1997-2011)
Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using 20 food groups
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include wave dummies.*

	(1)	(2)	(3)	(4)
	Dietary Diversity Score			
	Waves 1997-2011			
log(HH total income)	0.061***	0.060***	0.060***	0.059***
	[18.923]	[18.791]	[18.499]	[18.380]
Education	0.002***	0.002***	0.002***	0.002***
	[5.049]	[4.787]	[5.029]	[4.767]
HH size	-0.007***	-0.007***	-0.008***	-0.007***
	[-7.007]	[-6.539]	[-7.481]	[-7.005]
EDR [%]	-0.038***		-0.036***	
	[-6.838]		[-6.535]	
Age	0.028*	0.029*	0.336***	0.326***
	[1.810]	[1.893]	[3.225]	[3.134]
Gender	-0.005	-0.005	-0.005	-0.005
	[-1.003]	[-1.055]	[-0.961]	[-1.018]
Han nationality	-0.008**	-0.011***	-0.007	-0.009**
	[-1.962]	[-2.707]	[-1.629]	[-2.398]
Old home food habits	0.004	0.005	0.003	0.005
	[0.482]	[0.717]	[0.463]	[0.695]
Marital Status	0.034***	0.034***	0.034***	0.033***
	[5.478]	[5.340]	[5.416]	[5.282]
Urban	0.027***	0.028***	0.027***	0.028***
	[5.563]	[5.690]	[5.535]	[5.664]
EDR [diff]		-0.003***		-0.003***
		[-8.702]		[-8.435]
Age^2 (Head)			-0.280***	-0.270***
			[-2.962]	[-2.859]
Observations	5,748	5,748	5,748	5,748
R-squared	0.180	0.184	0.181	0.185
Wave Dummy	Y	Y	Y	Y

Table A.1. 14: Effects of interaction famine of (EDR [%/diff]) and Age/Age² on Dietary Diversity Score; Pooled regressions (1997-2011)
Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using 20 food groups
Robust standard errors are adjusted for heteroscedasticity; *significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively*; *t-statistics in brackets; regressions (1)-(4) include wave dummies, while regressions (5)-(8) include both wave and province dummies.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Dietary Diversity Score				
log(HH total income)	0.061***	0.060***	0.060***	0.059***	0.051***	0.051***	0.050***	0.050***
	[18.932]	[18.780]	[18.558]	[18.417]	[16.445]	[16.427]	[16.073]	[16.065]
Education	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	[5.054]	[4.751]	[5.036]	[4.746]	[4.889]	[4.864]	[4.878]	[4.865]
HH size	-0.007***	-0.007***	-0.008***	-0.007***	-0.003**	-0.002**	-0.003***	-0.003***
	[-7.016]	[-6.536]	[-7.408]	[-6.934]	[-2.542]	[-2.529]	[-2.993]	[-2.980]
EDR	0.037		-0.008					
	[1.575]		[-0.641]					
Age	0.262***	0.099***	0.296***	0.293***	0.122	0.048	0.249**	0.246**
	[3.498]	[3.353]	[2.781]	[2.750]	[1.628]	[1.634]	[2.376]	[2.353]
EDR [%] x Age	-0.134***				-0.068*			
	[-3.241]				[-1.636]			
Gender	-0.006	-0.006	-0.005	-0.006	-0.008*	-0.009*	-0.008	-0.008
	[-1.135]	[-1.176]	[-1.052]	[-1.097]	[-1.706]	[-1.715]	[-1.628]	[-1.635]
Han nationality	-0.008**	-0.011***	-0.007*	-0.010**	0.007*	0.007*	0.009**	0.009**
	[-2.070]	[-2.845]	[-1.704]	[-2.491]	[1.767]	[1.778]	[2.036]	[2.041]
Old home food habits	0.004	0.005	0.003	0.005	0.013*	0.013*	0.013*	0.013*
	[0.495]	[0.723]	[0.469]	[0.697]	[1.867]	[1.870]	[1.843]	[1.844]
Marital Status	0.033***	0.032***	0.033***	0.032***	0.031***	0.031***	0.031***	0.031***
	[5.252]	[5.122]	[5.257]	[5.137]	[5.044]	[5.022]	[5.052]	[5.036]
Urban	0.026***	0.027***	0.026***	0.027***	0.040***	0.040***	0.040***	0.040***
	[5.306]	[5.457]	[5.331]	[5.489]	[8.022]	[8.013]	[8.034]	[8.025]
EDR [diff]		0.001		-0.001*				
		[0.835]		[-1.726]				
EDR [diff] x Age		-0.008***				-0.005*		
		[-2.890]				[-1.833]		
Age²			-0.090	-0.196*			-0.166	-0.198*
			[-0.689]	[-1.881]			[-1.295]	[-1.930]
EDR x Age²			-0.088**				-0.033	
			[-2.270]				[-0.845]	
EDR [diff] x Age²				-0.005*				-0.003
				[-1.891]				[-1.014]
Observations	5,748	5,748	5,748	5,748	5,748	5,748	5,748	5,748
R-squared	0.181	0.185	0.182	0.185	0.246	0.246	0.247	0.247
Wave Dummy	Y	Y	Y	Y	Y	Y	Y	Y
Province Dummy					Y	Y	Y	Y
AIC	-10,234	-10,258	-10,235	-10,260	-10,697	-10,698	-10,700	-10,700
BIC	-10,120	-10,145	-10,115	-10,140	-10,544	-10,545	-10,540	-10,540

Table A.1. 15: Effects of famine (EDR [%/diff]) on Dietary Diversity Score; Pooled regressions (1997-2011)
Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using ***6 food groups***
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include wave dummies.*

	(1)	(2)	(3)	(4)
	Dietary Diversity Score			
log (HH total income)	0.077***	0.076***	0.076***	0.075***
	[16.504]	[16.345]	[16.241]	[16.095]
Education	0.005***	0.005***	0.005***	0.005***
	[6.947]	[6.678]	[6.932]	[6.664]
HH size	-0.010***	-0.010***	-0.011***	-0.010***
	[-6.603]	[-6.161]	[-6.784]	[-6.331]
EDR [%]	-0.060***		-0.059***	
	[-6.915]		[-6.747]	
Age	0.036	0.038	0.293*	0.278*
	[1.481]	[1.566]	[1.758]	[1.671]
Gender	-0.021**	-0.021**	-0.020**	-0.021**
	[-2.473]	[-2.523]	[-2.450]	[-2.503]
Han nationality	-0.001	-0.005	0.000	-0.004
	[-0.121]	[-0.785]	[0.043]	[-0.637]
Old home food habits	-0.001	0.002	-0.001	0.002
	[-0.071]	[0.156]	[-0.083]	[0.144]
Marital Status	0.057***	0.055***	0.056***	0.055***
	[5.678]	[5.553]	[5.633]	[5.511]
Urban	0.024***	0.025***	0.024***	0.025***
	[3.486]	[3.619]	[3.471]	[3.606]
EDR [diff]		-0.005***		-0.005***
		[-8.801]		[-8.654]
Age^2			-0.233	-0.218
			[-1.563]	[-1.464]
Observations	5,748	5,748	5,748	5,748
R-squared	0.137	0.141	0.137	0.142
AIC	-5,274	-5,302	-5,275	-5,303
BIC	-5,168	-5,196	-5,162	-5,190
Wave Dummy	Y	Y	Y	Y

Table A.1. 16: Effects of interaction famine of (EDR [%/diff]) and Age/Age² on Dietary Diversity Score; Pooled regressions (1997-2011)

*Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using **6 food groups***

Robust standard errors are adjusted for heteroscedasticity; *significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively*; *t-statistics in brackets; regressions (1)-(4) include wave dummies, while regressions (5)-(8) include both wave and province dummies.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dietary Diversity Score							
log (HH total income)	0.077***	0.076***	0.076***	0.075***	0.066***	0.066***	0.066***	0.066***
	[16.535]	[16.355]	[16.331]	[16.166]	[14.595]	[14.585]	[14.400]	[14.402]
Education	0.005***	0.005***	0.005***	0.005***	0.004***	0.004***	0.004***	0.004***
	[6.957]	[6.635]	[6.947]	[6.633]	[6.101]	[6.075]	[6.094]	[6.075]
HH size	-0.010***	-0.010***	-0.011***	-0.010***	-0.005***	-0.005***	-0.006***	-0.006***
	[-6.619]	[-6.162]	[-6.696]	[-6.230]	[-3.371]	[-3.360]	[-3.507]	[-3.488]
EDR [%]	0.072**		0.001					
	[2.077]		[0.073]					
Age	0.454***	0.178***	0.206	0.191	0.162	0.063	0.154	0.145
	[4.126]	[4.050]	[1.217]	[1.126]	[1.482]	[1.446]	[0.934]	[0.877]
EDR [%] x Age	-0.239***				-0.098			
	[-3.943]				[-1.628]			
Gender	-0.022***	-0.022***	-0.021***	-0.022***	-0.017**	-0.017**	-0.016**	-0.017**
	[-2.626]	[-2.678]	[-2.579]	[-2.634]	[-2.065]	[-2.082]	[-2.025]	[-2.041]
Han nationality	-0.002	-0.006	-0.000	-0.005	-0.004	-0.004	-0.003	-0.003
	[-0.237]	[-0.955]	[-0.058]	[-0.787]	[-0.565]	[-0.555]	[-0.424]	[-0.417]
Old home food habits	-0.001	0.002	-0.001	0.002	0.018	0.018	0.018	0.018
	[-0.054]	[0.168]	[-0.072]	[0.151]	[1.546]	[1.548]	[1.533]	[1.535]
Marital Status	0.054***	0.053***	0.054***	0.053***	0.049***	0.049***	0.049***	0.049***
	[5.437]	[5.292]	[5.438]	[5.302]	[5.194]	[5.156]	[5.195]	[5.163]
Urban	0.022***	0.023***	0.022***	0.023***	0.037***	0.037***	0.037***	0.037***
	[3.181]	[3.310]	[3.183]	[3.314]	[5.399]	[5.367]	[5.404]	[5.372]
EDR [diff]		0.004*		-0.001				
		[1.702]		[-0.673]				
EDR [diff] x Age		-0.016***				-0.008**		
		[-3.987]				[-2.054]		
Age²			0.179	-0.026			-0.034	-0.088
			[0.912]	[-0.162]			[-0.176]	[-0.560]
EDR [%] x Age²			-0.190***				-0.066	
			[-3.442]				[-1.194]	
EDR [diff] x Age²				-0.013***				-0.006
				[-3.468]				[-1.602]
Observations	5,748	5,748	5,748	5,748	5,748	5,748	5,748	5,748
R-squared	0.139	0.143	0.139	0.143	0.205	0.205	0.205	0.205
AIC	-5,286	-5,315	-5,284	-5,312	-5,733	-5,735	-5,732	-5,733
BIC	-5,173	-5,202	-5,164	-5,192	-5,580	-5,582	-5,572	-5,573
Wave Dummy	Y	Y	Y	Y	Y	Y	Y	Y
Province Dummy					Y	Y	Y	Y

Table A.1. 17: Effects of famine (EDR [%/diff]) on Dietary Diversity Score; Pooled regressions (1997-2011)
*Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using **12 food groups***
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include wave dummies.*

	(1)	(2)	(3)	(4)
	Dietary Diversity Score			
log (HH total income)	0.070***	0.069***	0.069***	0.068***
	[17.936]	[17.793]	[17.603]	[17.478]
Education	0.003***	0.003***	0.003***	0.003***
	[4.804]	[4.518]	[4.788]	[4.502]
HH size	-0.009***	-0.009***	-0.010***	-0.009***
	[-7.011]	[-6.537]	[-7.343]	[-6.853]
EDR [%]	-0.033***		-0.032***	
	[-4.679]		[-4.452]	
Age	0.024	0.025	0.333**	0.315**
	[1.211]	[1.275]	[2.438]	[2.306]
Gender	-0.007	-0.008	-0.007	-0.007
	[-1.031]	[-1.119]	[-0.999]	[-1.092]
Han nationality	-0.012**	-0.016***	-0.010*	-0.015***
	[-2.210]	[-3.254]	[-1.955]	[-3.023]
Old home food habits	0.008	0.010	0.008	0.010
	[0.941]	[1.170]	[0.926]	[1.154]
Marital Status	0.046***	0.045***	0.045***	0.044***
	[5.497]	[5.356]	[5.448]	[5.311]
Urban	0.025***	0.025***	0.025***	0.025***
	[4.074]	[4.214]	[4.053]	[4.195]
EDR [diff]		-0.003***		-0.003***
		[-6.821]		[-6.624]
Age^2			-0.280**	-0.263**
			[-2.281]	[-2.139]
Observations	5,748	5,748	5,748	5,748
R-squared	0.139	0.143	0.140	0.143
AIC	-7,481	-7,504	-7,484	-7,507
BIC	-7,374	-7,398	-7,371	-7,394
Wave Dummy	Y	Y	Y	Y

Table A.1. 18: Effects of interaction famine of (EDR [%/diff]) and Age/Age² on Dietary Diversity Score; Pooled regressions (1997-2011)

*Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using **12 food groups***

Robust standard errors are adjusted for heteroscedasticity; *significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively*; *t-statistics in brackets; regressions (1)-(4) include wave dummies, while regressions (5)-(8) include both wave and province dummies.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dietary Diversity Score							
log (HH total income)	0.070***	0.069***	0.069***	0.068***	0.058***	0.058***	0.057***	0.057***
	[17.953]	[17.793]	[17.660]	[17.513]	[15.560]	[15.549]	[15.258]	[15.256]
Education	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	[4.808]	[4.486]	[4.794]	[4.484]	[4.674]	[4.654]	[4.665]	[4.655]
HH size	-0.009***	-0.009***	-0.010***	-0.009***	-0.003**	-0.003**	-0.004***	-0.004***
	[-7.017]	[-6.532]	[-7.277]	[-6.792]	[-2.387]	[-2.377]	[-2.736]	[-2.727]
EDR [%]	0.051*		0.001					
	[1.715]		[0.039]					
Age	0.291***	0.101***	0.286**	0.278**	0.118	0.043	0.253*	0.251*
	[3.070]	[2.707]	[2.054]	[1.995]	[1.256]	[1.156]	[1.855]	[1.838]
EDR [%] x Age	-0.153***				-0.069			
	[-2.909]				[-1.326]			
Gender	-0.008	-0.008	-0.007	-0.008	-0.011*	-0.011*	-0.011	-0.011
	[-1.146]	[-1.219]	[-1.080]	[-1.157]	[-1.681]	[-1.688]	[-1.618]	[-1.623]
Han nationality	-0.012**	-0.017***	-0.011**	-0.016***	0.007	0.007	0.009	0.009
	[-2.303]	[-3.365]	[-2.022]	[-3.099]	[1.383]	[1.392]	[1.606]	[1.611]
Old home food habits	0.009	0.011	0.008	0.010	0.022**	0.022**	0.021**	0.021**
	[0.951]	[1.174]	[0.930]	[1.154]	[2.476]	[2.478]	[2.459]	[2.460]
Marital Status	0.044***	0.043***	0.044***	0.043***	0.038***	0.038***	0.038***	0.038***
	[5.308]	[5.182]	[5.314]	[5.197]	[4.815]	[4.797]	[4.821]	[4.808]
Urban	0.023***	0.024***	0.024***	0.024***	0.042***	0.042***	0.042***	0.042***
	[3.849]	[4.020]	[3.872]	[4.051]	[6.801]	[6.792]	[6.811]	[6.803]
EDR [diff]		0.002		-0.001				
		[0.804]		[-1.273]				
EDR [diff] x Age		-0.009**				-0.005		
		[-2.465]				[-1.483]		
Age²			-0.057	-0.182			-0.175	-0.207
			[-0.344]	[-1.351]			[-1.074]	[-1.576]
EDR [%] x Age²			-0.103**				-0.033	
			[-2.128]				[-0.683]	
EDR [diff] x Age²				-0.005*				-0.003
				[-1.654]				[-0.809]
Observations	5,748	5,748	5,748	5,748	5,748	5,748	5,748	5,748
R-squared	0.140	0.144	0.141	0.144	0.209	0.209	0.209	0.209
AIC	-7,487	-7,509	-7,486	-7,508	-7,951	-7,952	-7,952	-7,953
BIC	-7,374	-7,395	-7,367	-7,388	-7,798	-7,799	-7,793	-7,793
Wave Dummy	Y	Y	Y	Y	Y	Y	Y	Y
Province Dummy					Y	Y	Y	Y

Table A.1. 19: Effects of Famine (EDR [%/diff]) on Dietary Diversity Score of birth cohorts 1954-1962; Pooled *regressions (1997-2011)*
*Variables are defined as described in Table 1.2. EDR [%] represents Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR [diff] presents the difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using **20 food groups***
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include birth-cohort dummy (suppressed for brevity).*

	(1) Dietary Diversity Score	(2) Dietary Diversity Score
EDR [%]	-0.016	
	[-0.617]	
BirthYear1954 x EDR [%]	-0.032	
	[-0.984]	
BirthYear1955 x EDR [%]	-0.042	
	[-1.380]	
BirthYear1956 x EDR [%]	-0.012	
	[-0.371]	
BirthYear1957 x EDR [%]	-0.068**	
	[-1.985]	
BirthYear1958 x EDR [%]	-0.025	
	[-0.544]	
BirthYear1959 x EDR [%]	-0.007	
	[-0.183]	
BirthYear1960 x EDR [%]	0.034	
	[0.703]	
BirthYear1961 x EDR [%]	-0.007	
	[-0.155]	
BirthYear1962 x EDR [%]	0.097	
	[1.574]	
EDR[diff]		-0.002
		[-1.252]
BirthYear1954 x EDR [diff]		-0.002
		[-1.053]
BirthYear1955 x EDR [diff]		-0.002
		[-1.122]
BirthYear1956 x EDR [diff]		0.000
		[0.010]
BirthYear1957 x EDR [diff]		-0.004*
		[-1.809]
BirthYear1958 x EDR [diff]		-0.002
		[-0.577]
BirthYear1959 x EDR [diff]		0.000
		[0.072]
BirthYear1960 x EDR [diff]		0.003
		[0.940]
BirthYear1961 x EDR [diff]		-0.001
		[-0.492]
BirthYear1962 x EDR [diff]		0.006
		[1.420]
Observations	1,812	1,812
R-squared	0.195	0.199
Birth Cohort Dummy	Y	Y

Table A.1. 20: Effects of interaction of Famine inverted term (EDR_inv [diff]) and Income on Dietary Diversity Score; Cross sectional regressions

Variables are defined as described in Table 1.2. EDR_inv [diff] represents inverted value of Excess Death Rate calculated as difference between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); Dependent variable Dietary Diversity Score is calculated using 20 food groups.

*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; regressions (2), (4), (6) and (8) include province dummy variable.*

	Dietary Diversity Score							
	2004		2006		2009		2011	
log (HH total income)	0.047***	0.057***	0.039**	0.036**	0.022	0.017	0.039*	0.037*
	[3.386]	[4.028]	[2.241]	[2.015]	[1.596]	[1.262]	[1.844]	[1.729]
Education	0.001	0.001	0.003**	0.003**	0.002	0.002*	0.001	-0.000
	[1.005]	[1.253]	[2.423]	[2.398]	[1.394]	[1.740]	[0.421]	[-0.076]
EDR_inv [diff]	0.044**		0.073***		0.084***		0.036	
	[2.116]		[2.855]		[2.680]		[0.826]	
Age	0.012	-0.013	0.093**	0.056	0.007	-0.075	-0.105*	-0.139**
	[0.352]	[-0.360]	[2.034]	[1.238]	[0.128]	[-1.563]	[-1.874]	[-2.436]
EDR_inv [diff] x log (HH total income)	0.052**	0.013	0.043	0.037	0.034	0.022	0.041	0.038
	[2.088]	[0.526]	[1.425]	[1.291]	[1.408]	[0.959]	[1.262]	[1.125]
Gender	-0.002	-0.009	-0.014	-0.022	0.016	0.001	0.003	-0.001
	[-0.188]	[-0.873]	[-0.862]	[-1.340]	[0.970]	[0.067]	[0.215]	[-0.088]
HH size	-0.008***	-0.004*	-0.008***	-0.001	-0.004	0.004	-0.001	0.002
	[-3.532]	[-1.901]	[-2.682]	[-0.257]	[-1.231]	[1.633]	[-0.419]	[0.665]
Han nationality	-0.015	0.004	0.001	0.019**	-0.026***	0.013	-0.033***	-0.033***
	[-1.612]	[0.434]	[0.133]	[1.980]	[-2.631]	[1.262]	[-2.869]	[-2.605]
Old home food habits	0.008	0.010	0.028	0.043**	-0.013	-0.003	0.037	0.046*
	[0.439]	[0.539]	[1.567]	[2.392]	[-0.638]	[-0.145]	[1.399]	[1.747]
Marital Status	0.056***	0.059***	0.035*	0.023	0.010	0.013	0.032	0.031
	[4.859]	[4.305]	[1.962]	[1.272]	[0.449]	[0.620]	[1.473]	[1.437]
Urban	0.032***	0.045***	0.019	0.048***	0.052***	0.076***	0.035*	0.050**
	[2.942]	[4.191]	[1.182]	[2.943]	[2.625]	[3.837]	[1.730]	[2.407]
Observations	1,001	1,001	832	832	764	764	686	686
R-squared	0.171	0.233	0.187	0.292	0.144	0.294	0.188	0.225
Province Dummy		Y		Y		Y		Y

Table A.1. 21: Effects of interaction of Famine inverted term (EDR_inv [diff]) and Education on Dietary Diversity Score; Cross sectional regressions.

Variables are defined as described in Table 1.2. EDR_inv [diff] represents inverted value of Excess Death Rate, calculated as difference between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); Dependent variable Dietary Diversity Score is calculated using 20 food groups.

*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; regressions (2), (4), (6) and (8) include province dummy variable.*

	Dietary Diversity Score							
	2004		2006		2009		2011	
log (HH total income)	0.075***	0.064***	0.065***	0.058***	0.042***	0.030***	0.062***	0.058***
	[9.547]	[8.172]	[7.306]	[7.195]	[5.757]	[4.468]	[8.447]	[7.331]
Education	-0.001	-0.000	0.001	0.001	-0.001	-0.000	-0.000	-0.001
	[-0.612]	[-0.243]	[0.754]	[0.791]	[-0.268]	[-0.126]	[-0.141]	[-0.312]
EDR_inv [diff]	0.055**		0.092***		0.099***		0.083***	
	[2.506]		[4.231]		[3.591]		[2.634]	
Age	0.012	-0.014	0.097**	0.058	0.011	-0.073	-0.106*	-0.142**
	[0.340]	[-0.392]	[2.117]	[1.296]	[0.205]	[-1.525]	[-1.873]	[-2.451]
EDR_inv[diff] x Education	0.004	0.003	0.003	0.002	0.004	0.004	0.002	0.002
	[1.300]	[1.073]	[0.840]	[0.719]	[1.164]	[1.151]	[0.407]	[0.330]
Gender	-0.001	-0.009	-0.013	-0.021	0.020	0.004	0.003	-0.001
	[-0.111]	[-0.811]	[-0.816]	[-1.310]	[1.196]	[0.269]	[0.232]	[-0.075]
HH size	-0.008***	-0.004*	-0.008***	-0.001	-0.004	0.004	-0.001	0.002
	[-3.527]	[-1.888]	[-2.651]	[-0.236]	[-1.225]	[1.630]	[-0.396]	[0.699]
Han nationality	-0.015	0.004	0.002	0.020**	-0.024**	0.014	-0.032***	-0.032**
	[-1.631]	[0.362]	[0.217]	[2.027]	[-2.412]	[1.358]	[-2.824]	[-2.506]
Old home food habits	0.009	0.010	0.028	0.043**	-0.013	-0.003	0.036	0.045*
	[0.485]	[0.563]	[1.580]	[2.391]	[-0.654]	[-0.146]	[1.365]	[1.709]
Marital Status	0.060***	0.059***	0.035*	0.023	0.010	0.013	0.029	0.029
	[4.776]	[4.356]	[1.911]	[1.250]	[0.440]	[0.625]	[1.323]	[1.307]
Urban	0.035***	0.047***	0.019	0.048***	0.052***	0.076***	0.034*	0.050**
	[3.246]	[4.339]	[1.200]	[2.955]	[2.604]	[3.850]	[1.691]	[2.380]
Observations	1,001	1,001	832	832	764	764	686	686
R-squared	0.169	0.233	0.186	0.291	0.144	0.295	0.186	0.223
Province Dummy		Y		Y		Y		Y

Table A.1. 22: Effects of interaction of Famine inverted term (EDR_inv [diff]) and Gender on Dietary Diversity Score; Cross sectional regressions

Variables are defined as described in Table 1.2. EDR_inv [diff] represents inverted value of Excess Death Rate calculated as difference between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); Dependent variable Dietary Diversity Score is calculated using 20 food groups.

*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; regressions (2), (4), (6) and (8) include province dummy variable.*

	Dietary Diversity Score							
	2004		2006		2009		2011	
log (HH total income)	0.076***	0.064***	0.065***	0.058***	0.041***	0.029***	0.063***	0.058***
	[9.556]	[8.179]	[7.378]	[7.241]	[5.644]	[4.336]	[8.455]	[7.307]
Education	0.001	0.001	0.003**	0.003**	0.002	0.002*	0.001	-0.000
	[1.011]	[1.272]	[2.458]	[2.437]	[1.509]	[1.831]	[0.460]	[-0.022]
EDR_inv[diff]	0.076**		0.031		0.041		0.093**	
	[2.273]		[0.802]		[0.788]		[2.002]	
Age	0.012	-0.012	0.095**	0.057	0.010	-0.073	-0.105*	-0.140**
	[0.331]	[-0.358]	[2.070]	[1.271]	[0.198]	[-1.531]	[-1.853]	[-2.436]
EDR_inv[diff] x Gender	0.003	-0.018	0.080**	0.056	0.086	0.082*	-0.001	-0.009
	[0.083]	[-0.566]	[1.996]	[1.426]	[1.645]	[1.947]	[-0.020]	[-0.187]
Gender	-0.003	-0.001	-0.050**	-0.046**	-0.026	-0.040*	0.003	0.003
	[-0.202]	[-0.035]	[-2.282]	[-2.042]	[-1.009]	[-1.719]	[0.139]	[0.103]
HH size	-0.008***	-0.004*	-0.008***	-0.001	-0.003	0.005*	-0.001	0.002
	[-3.524]	[-1.894]	[-2.609]	[-0.212]	[-1.201]	[1.658]	[-0.406]	[0.692]
Han nationality	-0.015	0.004	0.002	0.020**	-0.025**	0.014	-0.033***	-0.032**
	[-1.631]	[0.435]	[0.219]	[2.020]	[-2.451]	[1.383]	[-2.807]	[-2.496]
Old home food habits	0.008	0.010	0.026	0.042**	-0.014	-0.004	0.036	0.045*
	[0.456]	[0.541]	[1.488]	[2.319]	[-0.687]	[-0.213]	[1.354]	[1.694]
Marital Status	0.059***	0.057***	0.042**	0.028	0.017	0.020	0.029	0.028
	[4.648]	[4.117]	[2.168]	[1.406]	[0.705]	[0.923]	[1.294]	[1.254]
Urban	0.034***	0.046***	0.020	0.049***	0.052***	0.076***	0.034*	0.049**
	[3.164]	[4.254]	[1.243]	[2.975]	[2.613]	[3.849]	[1.672]	[2.366]
Observations	1,001	1,001	832	832	764	764	686	686
R-squared	0.168	0.233	0.188	0.292	0.144	0.295	0.186	0.222
Province Dummy		Y		Y		Y		Y

Table A.1. 23: Effects of interaction of Famine inverted term (EDR_inv [%/diff]) and Gender on Dietary Diversity Score; Pooled regression (2004-2011), split sample Cohort 1954-1962 and Cohort 1963-1988
Variables are defined as described in Table 1.2. EDR_inv [%] represents inverted value of Excess Death Rate calculated as ratio between 3 famine years (1959-1961) and 3 pre-famine years (1956-1958); EDR_inv [diff] represents inverted value of Excess Death Rate calculated as difference between 3 famine years and 3 pre-famine years. Dependent variable Dietary Diversity Score is calculated using 20 food groups.
*Robust standard errors are adjusted for heteroscedasticity; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; all regressions include wave dummies and regressions (2), (4), (6) and (8) include both wave and province dummy variable.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Born 1954-1962				Born 1963-1988			
log (HH total income)	0.054***	0.045***	0.054***	0.045***	0.032**	0.033***	0.032***	0.033***
	[7.376]	[6.177]	[7.484]	[6.171]	[2.584]	[2.828]	[2.619]	[2.828]
Education	0.002	0.002*	0.001	0.002*	0.002	0.004	0.003	0.004
	[1.310]	[1.876]	[1.209]	[1.860]	[0.772]	[1.647]	[1.116]	[1.647]
EDR_inv [%]	0.057				0.129			
	[0.424]				[0.915]			
Age (Head)	0.153	0.179	0.162	0.179	-0.997**	-0.860**	-1.054***	-0.860**
	[1.034]	[1.299]	[1.106]	[1.301]	[-2.539]	[-2.307]	[-2.716]	[-2.307]
EDR_inv [%] x Gender	0.054	-0.041			-0.020	-0.205		
	[0.393]	[-0.318]			[-0.142]	[-1.298]		
Gender	0.025	0.060	0.043	0.041	0.009	0.126	0.025	0.153
	[0.305]	[0.759]	[0.543]	[0.571]	[0.119]	[1.470]	[0.267]	[1.440]
HH size	-0.013***	-0.004	-0.012***	-0.004	-0.009**	-0.005	-0.008**	-0.005
	[-4.307]	[-1.098]	[-3.980]	[-1.086]	[-2.157]	[-1.216]	[-2.068]	[-1.216]
Han nationality	-0.012	0.016	-0.008	0.016	-0.018	0.002	-0.018	0.002
	[-1.224]	[1.400]	[-0.826]	[1.398]	[-1.052]	[0.105]	[-1.075]	[0.105]
Old home food habits	-0.038*	-0.020	-0.033	-0.020	-0.090	-0.069	-0.091	-0.069
	[-1.738]	[-1.011]	[-1.568]	[-1.013]	[-1.496]	[-1.376]	[-1.526]	[-1.376]
Marital Status	0.045**	0.005	0.042*	0.005	0.017	0.005	0.020	0.005
	[2.100]	[0.258]	[1.954]	[0.256]	[0.809]	[0.265]	[0.966]	[0.265]
Urban	0.036**	0.066***	0.038**	0.066***	0.009	0.050*	0.010	0.050*
	[2.434]	[4.300]	[2.536]	[4.304]	[0.384]	[1.824]	[0.444]	[1.824]
EDR_inv[diff]			0.088				0.150	
			[0.732]				[0.966]	
EDR_inv[diff] x Gender			0.021	-0.010			-0.038	-0.223
			[0.173]	[-0.086]			[-0.244]	[-1.298]
Wave dummies	Y	Y	Y	Y	Y	Y	Y	Y
Province dummies		Y		Y		Y		Y
Observations	918	918	918	918	360	360	360	360
R-squared	0.209	0.303	0.216	0.303	0.203	0.321	0.216	0.321
AIC	-1,528	-1,631	-1,535	-1,631	-632	-678	-638	-678
BIC	-1,456	-1,530	-1,463	-1,530	-577	-600	-583	-600

Annex 2

Table A.2 1: Effects of famine on absolute intake of calories and nutrients; Pooled regressions (2004 -2011)
Dependent variables: Kcal presents presents absolute intake of calories of household head per day; Carbs presents absolute intake of carbohydrates of household head per day (g/capita/day); Fat presents absolute intake of fat of household head per day (g/capita/day); Protein intake presents absolute intake of protein of household head per day (g/capita/day)
Explanatory variables: log(HH total income) presents logarithmic value of Household total income; Education, EDR [%], Age, Gender, HH Size, Han nationality, 'Old home' food habits, Marital Status Urban is defined as in Table 1.2
*Robust standard errors are adjusted for heteroscedasticity; standard errors are clustered by province; significance levels of 1%, 5%, and 10% are denoted by ***, **, and * respectively; t-statistics in brackets; all regressions include wave dummies.*

	(1) Kcal	(2) Carbs	(3) Fat	(4) Protein	(5) Kcal	(6) Carbs	(7) Fat	(8) Protein
log (HH total income)	64.101*	-6.011	5.442**	3.767**	60.708*	-6.386	5.434**	3.665**
	[2.329]	[-1.356]	[3.058]	[2.832]	[2.207]	[-1.428]	[3.029]	[2.759]
Education	-5.547	-2.461	0.801**	-0.124	-5.787	-2.487	0.801**	-0.131
	[-0.677]	[-1.700]	[3.197]	[-0.483]	[-0.688]	[-1.689]	[3.175]	[-0.504]
EDR [%]	-529.473**	-75.268**	-15.273*	-20.958**	-513.949**	-73.553**	-15.236*	-20.494**
	[-2.384]	[-2.046]	[-1.749]	[-2.215]	[-2.311]	[-1.963]	[-1.758]	[-2.166]
Age	-1,282.404***	-230.470***	-14.695*	-45.858***	1,933.703*	124.864	-6.964	50.386
	[-9.099]	[-7.572]	[-2.281]	[-7.334]	[2.020]	[0.578]	[-0.081]	[0.970]
Age^2					-2,760.348**	-304.979	-6.636	-82.605*
					[-3.411]	[-1.785]	[-0.090]	[-1.963]
Gender	418.888***	71.731***	0.565	11.923***	422.994***	72.185***	0.575	12.046***
	[5.753]	[7.556]	[0.150]	[7.718]	[5.893]	[7.696]	[0.151]	[8.064]
HH size	12.619	7.918	-2.144*	0.691	7.220	7.322	-2.157*	0.530
	[0.842]	[1.830]	[-2.092]	[1.051]	[0.505]	[1.624]	[-2.053]	[0.802]
Han nationality	-111.128	-24.895	1.800	-2.663	-99.712	-23.634	1.828	-2.322
	[-0.807]	[-1.203]	[0.337]	[-0.569]	[-0.740]	[-1.167]	[0.338]	[-0.504]
Old home food habits	80.469	11.382	4.812	1.230	80.065	11.337	4.811	1.218
	[1.432]	[0.877]	[1.465]	[0.715]	[1.414]	[0.864]	[1.463]	[0.655]
Marital Status	-109.751	-39.421*	4.943	-0.227	-116.929	-40.214*	4.926	-0.442
	[-1.070]	[-2.189]	[0.934]	[-0.078]	[-1.154]	[-2.259]	[0.942]	[-0.153]
Urban	-138.453	-26.240	-3.819	-2.825	-139.317	-26.335	-3.821	-2.851
	[-1.442]	[-1.025]	[-1.080]	[-0.739]	[-1.514]	[-1.052]	[-1.077]	[-0.755]
Wave dummies	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3,283	3,283	3,283	3,283	3,283	3,283	3,283	3,283
R-squared	0.118	0.123	0.063	0.107	0.120	0.124	0.063	0.108
AIC	52,338	40,632	33,132	30,607	52,332	40,631	33,132	30,605
BIC	52,381	40,674	33,174	30,650	52,375	40,680	33,174	30,654